

SEARCH REQUEST FORM  
Scientific and Technical Information Center

Access DB# \_\_\_\_\_

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
Art Unit: \_\_\_\_\_ Phone Number 301 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Mail Box and Bldg./Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

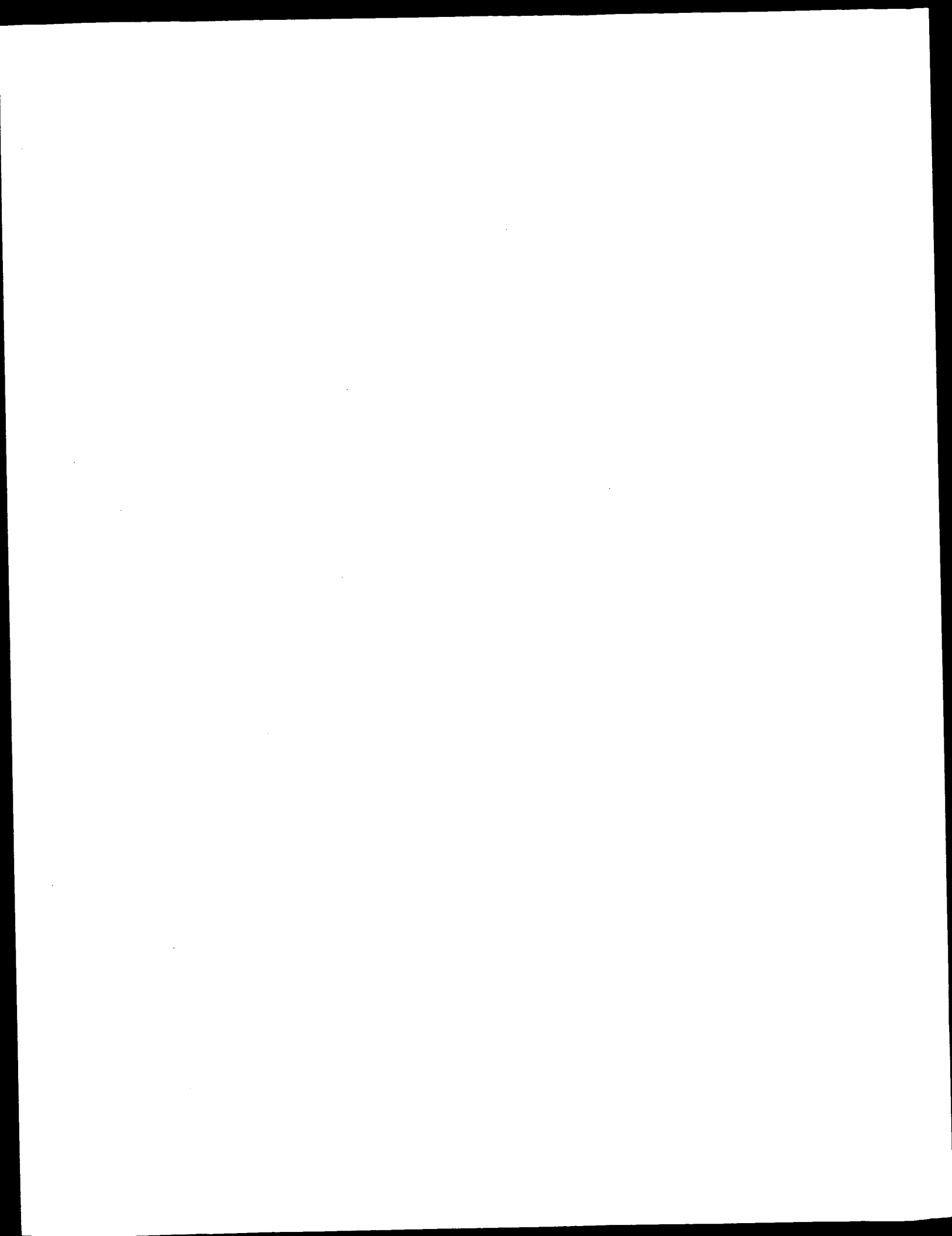
*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

|                                       | Type of Search            | Vendors and cost where applicable |
|---------------------------------------|---------------------------|-----------------------------------|
| Searcher <u>Toby Port</u>             | NA Sequence (#) <u>13</u> | STN _____                         |
| Searcher Phone # <u>308-3534</u>      | AA Sequence (#) _____     | Dialog _____                      |
| Searcher Location _____               | Structure (#) _____       | Questel Orbit _____               |
| Date Searcher Picked Up <u>10/24</u>  | Bibliographic _____       | On Line _____                     |
| Date Completed <u>10/25</u>           | Litigation _____          | Lexis Nexis _____                 |
| Searcher Prep & Review Time <u>15</u> | Fulltext _____            | Sequence Systems <u>CS</u>        |
| Client Prep Time _____                | Patent Family _____       | WWW Internet _____                |
| File Time <u>15</u>                   | Other _____               | Other Inquiries _____             |

PT 1000 10/25

BEST AVAILABLE COPY





|      |             |
|------|-------------|
| 1:   | gb_est1.*   |
| 2:   | gb_est2.*   |
| 3:   | gb_est3.*   |
| 4:   | gb_est4.*   |
| 5:   | gb_est5.*   |
| 6:   | gb_est6.*   |
| 7:   | gb_est7.*   |
| 8:   | gb_est8.*   |
| 9:   | gb_est9.*   |
| 10:  | gb_est10.*  |
| 11:  | gb_est11.*  |
| 12:  | gb_est12.*  |
| 13:  | gb_est13.*  |
| 14:  | gb_est14.*  |
| 15:  | gb_est15.*  |
| 16:  | gb_est16.*  |
| 17:  | gb_est17.*  |
| 18:  | gb_est18.*  |
| 19:  | gb_est19.*  |
| 20:  | gb_est20.*  |
| 21:  | gb_est21.*  |
| 22:  | gb_est22.*  |
| 23:  | gb_est23.*  |
| 24:  | gb_est24.*  |
| 25:  | gb_est25.*  |
| 26:  | gb_est26.*  |
| 27:  | gb_est27.*  |
| 28:  | gb_est28.*  |
| 29:  | gb_est29.*  |
| 30:  | gb_est30.*  |
| 31:  | gb_est31.*  |
| 32:  | gb_est32.*  |
| 33:  | gb_est33.*  |
| 34:  | gb_est34.*  |
| 35:  | gb_est35.*  |
| 36:  | gb_est36.*  |
| 37:  | gb_est37.*  |
| 38:  | gb_est38.*  |
| 39:  | gb_est39.*  |
| 40:  | gb_est40.*  |
| 41:  | gb_est41.*  |
| 42:  | gb_est42.*  |
| 43:  | gb_est43.*  |
| 44:  | gb_est44.*  |
| 45:  | gb_est45.*  |
| 46:  | gb_est46.*  |
| 47:  | gb_est47.*  |
| 48:  | gb_est48.*  |
| 49:  | gb_est49.*  |
| 50:  | gb_est50.*  |
| 51:  | gb_est51.*  |
| 52:  | gb_est52.*  |
| 53:  | gb_est53.*  |
| 54:  | gb_est54.*  |
| 55:  | gb_est55.*  |
| 56:  | gb_est56.*  |
| 57:  | gb_est57.*  |
| 58:  | gb_est58.*  |
| 59:  | gb_est59.*  |
| 60:  | gb_est60.*  |
| 61:  | gb_est61.*  |
| 62:  | gb_est62.*  |
| 63:  | gb_est63.*  |
| 64:  | gb_est64.*  |
| 65:  | gb_est65.*  |
| 66:  | gb_est66.*  |
| 67:  | gb_est67.*  |
| 68:  | gb_est68.*  |
| 69:  | gb_est69.*  |
| 70:  | gb_est70.*  |
| 71:  | gb_est71.*  |
| 72:  | gb_est72.*  |
| 73:  | gb_est73.*  |
| 74:  | gb_est74.*  |
| 75:  | gb_est75.*  |
| 76:  | gb_est76.*  |
| 77:  | gb_est77.*  |
| 78:  | gb_est78.*  |
| 79:  | gb_est79.*  |
| 80:  | gb_est80.*  |
| 81:  | gb_est81.*  |
| 82:  | gb_est82.*  |
| 83:  | gb_est83.*  |
| 84:  | gb_est84.*  |
| 85:  | gb_est85.*  |
| 86:  | gb_est86.*  |
| 87:  | gb_est87.*  |
| 88:  | gb_est88.*  |
| 89:  | gb_est89.*  |
| 90:  | gb_est90.*  |
| 91:  | gb_est91.*  |
| 92:  | gb_est92.*  |
| 93:  | gb_est93.*  |
| 94:  | gb_est94.*  |
| 95:  | gb_est95.*  |
| 96:  | gb_est96.*  |
| 97:  | gb_est97.*  |
| 98:  | gb_est98.*  |
| 99:  | gb_est99.*  |
| 100: | gb_est100.* |

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



|                       | 130 a   | 56 c            | 101 g | 120 t          | 6 others          |
|-----------------------|---|-----------------|-------|----------------|-------------------|
| ORIGIN                |   |                 |       |                |                   |
| Query Match           | 3.0%;   | Score 38.2;     | DB 3; | Length 413;    |                   |
| Best Local Similarity | 54.7%;  | Pred. NO. 0.84; |       |                |                   |
| Matches               | 76;   | Conservative    | 0;    | Mismatches 63; | Indels 0; Gaps 0; |
| 483                   | cccaagcggataccgaagatgcgtgtacacgagcttcggtggaatggacaagatgggc      | 542             |       |                |                   |
| 20y                   |   |                 |       |                |                   |
| 140                   | CTCTGGATGATGAGACAGAGAGTCTCTTAGGAAGCTTCAC                        | 199             |       |                |                   |
| Db                    |   |                 |       |                |                   |
| 543                   | cgctgaaaaatcgttcaccatttcacatatcacaaattgcagctgaagtcgtctcgcat     | 602             |       |                |                   |
| 20y                   |   |                 |       |                |                   |
| 200                   | TTTGAAAAGGCTTAACCAAGGATTCAGGTCGCCGGTTTGAGTCTTGAGAAATTTATTTGCTTT | 259             |       |                |                   |
| Db                    |   |                 |       |                |                   |
| 603                   | cggagagccagcggagcat   | 621             |       |                |                   |
| 20y                   |   |                 |       |                |                   |
| Db                    |   |                 |       |                |                   |
| 260                   | TGAGCATCTCTGAAGAAGTT  | 278             |       |                |                   |
| Db                    |   |                 |       |                |                   |

|            |  |
|------------|--|
| RESULT     | 3  |
| AA1516896  |  |
| LOCUS      | AA1516896 516 bp mRNA EST 14-JUL-1997  |
| DEFINITION | vhs7e9y.r1 Barstead mouse myotubes MP1RB5 Mus musculus cDNA clone IMAGE:902248 5' similar to SW:RS9_HUMAN P46781 40S RIBOSOMAL |

PROTEIN S9. ;, mRNA sequence.  
AA516896  
AA516896.1 GI:2256355  
EST.  
EST.  
mouse mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 516)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubouque, T.,  
Garrison, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:522912  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 473.  
 Location/Qualifiers  
     1..516  
         /organism="Mus musculus"  
         /strain="C3H"  
         /db\_xref="taxon:10090"



```

BASE COUNT      171 a      107 c      87 g      113 t      20 others
ORIGIN

Query Match
Best Local Similarity 2.9%; Score 36.8; DB 231; Length 498;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 832 gtaagaacgggcatcagttcggacaaatacgaaccccttagttatcttggttcgac 891
Db 271 GTTCTGGCGGCAATCAGGCAAGACAAAGAAATAAAGGTTAATTAAGTTAGAAAGAG 330

Qy 892 catgtgatctactgtatgttgcgaatgtcctgcctgattattgaaatcagcaggac 951
Db 331 GAAGTCAAAATTCCTCTGTTGAGATGACATGTTTATTAGAAATCCCATCGAC 390

RESULT 8
AW818171 518 bp mRNA EST 17-MAY-2000
LOCUS CM1-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW818171
ACCESSION AW818171
VERSION AW818171.1 GI:7911165
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 518)
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., M.R.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-CM1-ST0277-161
299-070-h05&t3=1999-12-16&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 76
High quality sequence stop: 183.

FEATURES
source
1..518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0277"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      193 a      119 c      89 g      117 t
ORIGIN

Query Match
Best Local Similarity 2.9%; Score 36.8; DB 121; Length 518;
Matches 61.5%; Pred. No. 2.5;

Qy 952 agaataaaactgtggaatatttaagtagtctatcattacacaaataaccgccgcg 1011
Db 1469 AAAAAAAMRYKWKRRKGRKRKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYK 1528

Qy 1012 cccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 1071
Db 1529 WKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 1588

Qy 1072 ctaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 1131
Db 1589 GTWGSMSGCGYKWKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYK 1648

Qy 1132 attccattgga 1143
Db 1649 WTTCTMTRKGA 1660

RESULT 7
AQ615572 498 bp DNA GSS 15-JUN-1999
LOCUS HS_5144_B1_B02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=720 Col=3 Row=D, DNA sequence.
ACCESSION AQ615572
VERSION AQ615572.1 GI:5076848
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 498)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end web Server:
http://www.htsc.washington.edu
Plate: 720 row: D column: 3
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 498.

FEATURES
source
1..498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

```

Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 755 agaccacaaacgaatttaactctcgaatgtaccaggtgttaatttagaagatttaatt 814  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 168 AGGATTAAACCAATTTGATCTCATGTGACACCAAGCAAAATATAGTAAAGATT 227  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 815 atgcctgttagaagtgttaagaacaggcgattca 850  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 228 CTGCTCAGAGTAACGTCTTAACACCTGCCCAATCA 263

RESULT 9  
 A1816449/c  
 LOCUS A1816449 431 bp mRNA EST 09-JUL-1999  
 DEFINITION au46f05.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
 IMAGE:2517825 5' similar to TR:Q19737 Q19737 F2E10.5 ; , mRNA  
 sequence.  
 ACCESSION A1816449  
 VERSION A1816449.1 GI:5431995  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 431)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,  
 J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
 White,Y., Wyllie,T., Waterston,R. and Wilson,R.  
 TITLE WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 407.

FEATURES  
 source  
 1..431  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2517825"  
 /clone\_lib="Schneider fetal brain 00004"  
 /sex="male"  
 /tissue\_type="frontal lobe"  
 /dev\_stage="5 months post-conception"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);  
 Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was  
 prepared from human fetal brain tissue. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence:  
 5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCCGCCCCCCCC-3'  
 and 3' adaptor sequence:  
 5'-GAGAGAGAGAGAGCTCGAGTTTCTTTTCTTTT-3'. The library was  
 size-selected for >0.5 kb inserts and has an average  
 insert size estimated at 1.2 kb. This library was  
 constructed using the CAP-trapper method for full-length  
 enrichment and has not undergone amplification. Library  
 was constructed by Dr. Claudio Schneider (LNCIB-Area  
 Science Park, Trieste, Italy)."  
 BASE COUNT 44 a 172 c 159 g 56 t  
 ORIGIN

Query Match 2.8%; Score 36.4; DB 102; Length 431;  
 Best Local Similarity 59.8%; Pred. No. 3.2;  
 Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 40 agccgcggggctgattaccgccggctccccaacctctgctaaccccgcttgcta 99  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 346 AGCAGCCAGGTCCAGTAGCTGCAGCGCGGCTCGAGCAGCAGAGCGCCCGCGCTG 287  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 100 taatgggttctccattccgctggtgttcaactctgaac 141  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 286 TAGCGGTGCTCTCCAGTCGCCGAGCTGCGCGCGCTCAGC 245

RESULT 10  
 A1879137/c  
 LOCUS A1879137 459 bp mRNA EST 23-AUG-1999  
 DEFINITION au55a12.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
 IMAGE:2518654 5' similar to TR:Q19737 Q19737 F2E10.5 ; , mRNA  
 sequence.  
 ACCESSION A1879137  
 VERSION A1879137.1 GI:5553186  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 459)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,  
 J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
 White,Y., Wyllie,T., Waterston,R. and Wilson,R.  
 TITLE WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 438.

FEATURES  
 source  
 1..459  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2518654"  
 /clone\_lib="Schneider fetal brain 00004"  
 /sex="male"  
 /tissue\_type="frontal lobe"  
 /dev\_stage="5 months post-conception"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);  
 Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was  
 prepared from human fetal brain tissue. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence:  
 5'-CAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCCGCCCCCCCC-3'  
 and 3' adaptor sequence:  
 5'-CAGAGAGAGAGAGCTCGAGTTTCTTTTCTTTT-3'. The library was  
 size-selected for >0.5 kb inserts and has an average  
 insert size estimated at 1.2 kb. This library was  
 constructed using the CAP-trapper method for full-length  
 enrichment and has not undergone amplification. Library  
 was constructed by Dr. Claudio Schneider (LNCIB-Area  
 Science Park, Trieste, Italy)."  
 BASE COUNT 78 a 146 c 138 g 97 t  
 ORIGIN

Query Match 2.8%; Score 36.4; DB 103; Length 459;  
 Best Local Similarity 59.8%; Pred. No. 3.2;  
 Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 40 agccgcggggctgattaccgccggctccccaacctctgctaaccccgcttgcta 99  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

```

RESULT 12
GC499317/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 914)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLCMI477 row: j column: 21
High quality sequence stop: 684.
Location/Qualifiers
1. 914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/30255="IMAGE:4668020"
FEATURES
source

```

```

/clone_lib="NIH_MGC_60"
/clone_lib="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (11 phage-resistant)"
/notes="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggcgcctggcc); Site_2: SfII (ggccattggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGGCGCGGACATG-dT(30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT      185 a      233 c      255 g      241 t
ORIGIN

Query Match      2.8%; Score 36.4; DB 154; Length 914;
Best Local Similarity 59.8%; Pred. No. 4;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0

QY 40 agcgcgggggtagtattacccgcggcgtccccaaacctgtgtaaccccggttgcta 99
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 AGCAGCCAGGTCCAGTAGAGTGCAGCGCGCGTCTGAGCAGCAGACGCCGCGCGCTG 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 100 taatgggttctccattggcgcgtggtgtcttcactctgaac 141
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 TAGCGGTGCTCTCCAGTCGCCGCGCAGCTCGCGCGGCTCAGC 78
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AZ444169/c
LOCUS
DEFINITION
ACCESSION
VERSION

```

## KEYWORDS



|                       |  |
|-----------------------|--|
| DEFINITION            | uz35d05.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3671049 5' similar to SW:RS2_MOUSE P25444 40S RIBOSOMAL PROTEIN S2 ;, mRNA sequence.  |
| ACCESSION             | BF152590   |
| VERSION               | BF152590.1   |
| KEYWORDS              | EST.   |
| SOURCE                | house mouse.   |
| ORGANISM              | Mus musculus   |
| REFERENCE             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 678)  |
| AUTHORS               | NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .   |
| TITLE                 | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  |
| JOURNAL               | Unpublished (1997)   |
| COMMENT               | Contact: Robert Strausberg, Ph.D.<br>Email: cgapbs-r@mail.nih.gov<br>Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys<br>cDNA Library Preparation: Life Technologies, Inc.<br>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)<br>DNA Sequencing by: Washington University Genome Sequencing Center<br>Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov/image/html/iresources.shtml">image.llnl.gov/image/html/iresources.shtml</a> |
| FEATURES              | <p>source</p> <p>1..678</p> <p>/organism="Mus musculus"</p> <p>/strain="C57/B6"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:3671049"</p> <p>/clone_lib="NCI_CGAP_Mam5"</p> <p>/tissue_type="tumor, gross tissue"</p> <p>/dev_stage="7 months"</p> <p>/lab_host="DH10B"</p> <p>/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"</p>                             |
| BASE COUNT            | 127 a 171 c 234 g 146 t  |
| ORIGIN                |  |
| Query Match           | 2.8%; Score 35.8; DB 145; Length 678;  |
| Best Local Similarity | 53.1%; Pred. No.5.7;   |
| Matches               | 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;   |
| Y                     | 134 ctctgaactacgacagacagaggcggcggaattgttaaggcgcatcgagtcacctta 193  |
| b                     | 566 CTGTCACTTCGATGGAACAGTGTGGGGTGCCATCTTGTTCCCCCACTAGCCTCTCCGC 507   |
| Y                     | 194 acttggtctatctagtattggggcagaggttgccacttcaattggccaacgacacctcc 253  |
| b                     | 506 ACAGGACCATGGAAGAGCTTGCCCAAGATGATGGCCCTCGTATGGCAGTAGCAACCTCC 447  |
| Y                     | 254 aaggattcaccacttgaagacc 276   |
| b                     | 446 TTGGACCACTTAACACCAGACCC 424  |
| RESULT 15             |  |
| Q089903/c             |  |
| OCUS                  |  |
| DEFINITION            | AQ089903 318 bp DNA GSS 26-AUG-1998  |
| REFERENCE             | HS_3001_A2_H04_WR C1T Approved Human Genomic Sporn Library D Homo sapiens genomic clone Plate-3001 Col=8 Row=O, DNA sequence.  |
| ACCESSION             | AQ089903   |
| VERSION               | AQ089903.1   |
| KEYWORDS              | GI:3458814   |
| SOURCE                | GSS.   |
| human                 |  |



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:19 ; Search time 180.6 Seconds  
(without alignments)  
818.671 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_211\_991  
Perfect score: 781  
Sequence: 1 gttggggacgaggtgcacc.....ttaaagtatgtcatctaa 781

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description       |
|------------|-------|---------------|--------|-------|-------------------|
| 1          | 162.6 | 20.8          | 1096   | 2     | US-08-418-071-4   |
| 2          | 142.6 | 18.3          | 1091   | 2     | US-08-418-071-5   |
| 3          | 137.8 | 17.6          | 1106   | 2     | US-08-418-071-3   |
| 4          | 108   | 13.8          | 1022   | 4     | US-08-793-634B-2  |
| 5          | 99.2  | 12.7          | 1017   | 4     | US-08-793-634B-6  |
| 6          | 62.8  | 8.0           | 1110   | 1     | US-08-202-186-14  |
| 7          | 59.6  | 7.6           | 982    | 3     | US-08-973-068-28  |
| 8          | 59.6  | 7.6           | 1110   | 1     | US-08-202-186-11  |
| 9          | 59.6  | 7.6           | 1111   | 1     | US-08-202-186-9   |
| 10         | 59.6  | 7.6           | 1111   | 1     | US-08-202-186-12  |
| 11         | 59.6  | 7.6           | 1111   | 1     | US-08-202-186-24  |
| 12         | 58.6  | 7.5           | 1103   | 1     | US-08-202-186-16  |
| 13         | 58.6  | 7.5           | 1103   | 1     | US-08-202-186-18  |
| 14         | 58    | 7.4           | 1109   | 1     | US-08-202-186-13  |
| 15         | 57    | 7.4           | 1111   | 1     | US-08-202-186-15  |
| 16         | 57    | 7.3           | 1104   | 1     | US-08-202-186-17  |
| 17         | 56.4  | 7.2           | 1111   | 1     | US-08-202-186-10  |
| 18         | 46.4  | 5.9           | 287    | 2     | US-08-418-071-1   |
| 19         | 46.4  | 5.9           | 300    | 2     | US-08-418-071-2   |
| 20         | 37.2  | 4.8           | 7218   | 1     | US-08-232-463-14  |
| 21         | 31.8  | 4.1           | 498    | 1     | US-07-781-254A-14 |
| 22         | 31.8  | 4.1           | 1338   | 1     | US-07-781-254A-6  |
| 23         | 31.8  | 4.1           | 1338   | 1     | US-07-781-254A-7  |
| 24         | 31.8  | 4.1           | 1863   | 1     | US-07-781-254A-4  |
| 25         | 31.8  | 4.1           | 1863   | 1     | US-07-781-254A-5  |
| 26         | 31    | 4.0           | 1212   | 3     | US-09-046-578-5   |
| 27         | 30.4  | 3.9           | 1212   | 3     | US-09-046-578-1   |

|      |      |     |       |   |                   |                    |
|------|------|-----|-------|---|-------------------|--------------------|
| 28   | 30   | 3.8 | 1437  | 3 | US-08-724-814-15  | Sequence 15, Appl  |
| C 29 | 29.8 | 3.8 | 951   | 1 | US-08-671-525B-1  | Sequence 1, Appl   |
| C 30 | 29.8 | 3.8 | 951   | 1 | US-08-672-109B-1  | Sequence 1, Appl   |
| C 31 | 29.8 | 3.8 | 951   | 1 | US-08-842-045-1   | Sequence 1, Appl   |
| C 32 | 29.8 | 3.8 | 951   | 2 | US-08-842-238-1   | Sequence 1, Appl   |
| C 33 | 29.8 | 3.8 | 951   | 3 | US-08-629-335B-1  | Sequence 1, Appl   |
| C 34 | 29.8 | 3.8 | 1633  | 1 | US-07-866-979-5   | Sequence 5, Appl   |
| C 35 | 29.8 | 3.8 | 1633  | 3 | US-08-466-906B-5  | Sequence 5, Appl   |
| C 36 | 29.8 | 3.8 | 1633  | 2 | US-08-706-281A-5  | Sequence 5, Appl   |
| C 37 | 29.8 | 3.8 | 1633  | 4 | US-09-201-746-5   | Sequence 5, Appl   |
| C 38 | 29.4 | 3.8 | 8920  | 2 | US-08-446-855A-1  | Sequence 1, Appl   |
| C 39 | 29.4 | 3.8 | 8920  | 4 | US-09-150-741-1   | Sequence 1, Appl   |
| C 40 | 29.2 | 3.7 | 1209  | 6 | 5352575-4         | Patent No. 5352575 |
| C 41 | 29.2 | 3.7 | 1596  | 5 | PCT-US94-11328A-3 | Sequence 3, Appl   |
| C 42 | 28.8 | 3.7 | 10607 | 1 | US-08-078-090-3   | Sequence 3, Appl   |
| C 43 | 28.6 | 3.7 | 1593  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
| C 44 | 28.6 | 3.7 | 1593  | 2 | US-08-975-114A-2  | Sequence 2, Appl   |
| C 45 | 28.6 | 3.7 | 1593  | 3 | US-08-849-281A-2  | Sequence 2, Appl   |

## ALIGNMENTS

RESULT 1  
US-08-418-071-4  
; Sequence 4, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh  
; APPLICANT: You, Li-Ru  
; APPLICANT: Soong, Tai-Seng  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/418,071  
; FILING DATE: 06-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: DCB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1096 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; MOLECULE TYPE: genomic DNA  
; DESCRIPTION: /desc "BTV DNA II (clone 2-17)"  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Banana Bunchy Top Virus (BTV)  
US-08-418-071-4

| Query Match           | 20.8%   | Score 162.6   | DB 2      | Length 1096 |
|-----------------------|---|---|-----------|-------------|
| Best Local Similarity | 53.4%   | Pred. No. 2e-44   |           |             |
| Matches 417           | Conservative 0  | Mismatches 349  | Indels 15 | Gaps 3      |
| 1                     | gttgaggacgaggtgtcaccttcaactggccaacgacacacccctccaagagttatcatcacttg | 60  |           |             |
| 2                     |   |   |           |             |
| 3                     |   |   |           |             |
| 4                     | 173   | GTGCGCAGCAAGTCGCTCCGGCCACCGCGCAAGAACCCCTCCAGGATATCATCCCTG     | 232       |             |
| 5                     | 61  | aagaccggtcgcgactgcaagatgaagactgtcttgggaatcacaggattcaccttg     | 120       |             |
| 6                     |   |   |           |             |
| 7                     | 233   | AAAAAATCAATCGCTCCGGGATGAAAAAGAAATATGG--CTCCCGTCTCACTGG        | 289       |             |
| 8                     | 121   | gagcgaccggtgttccgacacagaatagactactgttcgaaggaacgaggtgctt       | 180       |             |
| 9                     |   |   |           |             |
| 10                    | 290   | GAGATTCCGAAGAAGTACGACAAACAGAAATCGAGATACTGTCGAAGGAAACCTAGTT    | 349       |             |
| 11                    | 181   | ctcgagcacgaggtcccgactcgtcctgagtcataaaggccacgattggcccaagatt    | 240       |             |
| 12                    |   |   |           |             |
| 13                    | 350   | CTTGAATCGGTATCTCCGGTGGTCTCGTTCGAAGAAGCGCAAGCTTCTCGATAGATC     | 409       |             |
| 14                    | 241   | gctgaggaacatgatgaactccgctgggaagaccgagcggtatcacgaaatgcgttga    | 300       |             |
| 15                    |   |   |           |             |
| 16                    | 410   | AGAGAGAGCCCTGAGGAATTTGAAGATGGACGATCCATCCAAGTATCGACAGATCTTGGCA | 469       |             |
| 17                    | 301   | cacgagcttcggtgggaatggaacaaatggtggcgtgaaatccgttcccatctccat     | 360       |             |
| 18                    |   |   |           |             |
| 19                    | 470   | GTGGAATCAATTAAGATGCCAAATTAATTCGGAATCGGTTACGAACTA-----         | 520       |             |
| 20                    | 361   | cacaattggcagcttgaagtgtcgtcgtcgatcgagagcgagcgagcatcgcaaatc     | 420       |             |
| 21                    | 521   | AAAGATGGCAAAATAAATTAATCAACACATCGAAGGTGTTCTGTGATCGAAGTATC      | 580       |             |
| 22                    | 421   | ctctggatatcgacagacgagcagggaggaactcgtgttgcgaatactctcgga        | 480       |             |
| 23                    | 581   | ATCTCGGTATACGGTCCCAACGGAGCGAAGAAAGTCAACCTTCGCAAGATATCTATCA    | 640       |             |
| 24                    | 481   | ctcaagcccgactggttctacacatgtgttggaaaccgaaagagcgattgtaccagtac   | 540       |             |
| 25                    | 641   | TTAAACCCGGATGGGATATATCAACGGTGGAAAGACGTCGGATATGATGCACATCAT     | 700       |             |
| 26                    | 541   | atcgaggaccacaacgaatttaactctgatgtaccagggtgaatttagagtattta      | 600       |             |
| 27                    | 701   | ACGATCGATCTGTATATCAATCTGGATTATTTGATATATCCCCGAAGTCATTCAGATTCTG | 760       |             |
| 28                    | 601   | aattatgccctgttagaatgtttaagaacagggcattcagtcggacaaatacgaacc-    | 659       |             |
| 29                    | 761   | AATTATGGCGTTATAGACAAATTAACAATAGAGTTTAAATAAATACAAAAATACGAACCA  | 820       |             |
| 30                    | 660   | --ccttagtatacttgggttcgacacatgtcagctactcgttatttgcgaagtctcgct   | 717       |             |
| 31                    | 821   | TGTTGATTATAGAAAAGATGGCAAAATGTCCATGTAAATTTATATGCGAAATGTGTGCCT  | 880       |             |
| 32                    | 718   | gattattgaaatcagcagggacagaataaactgtggaattatttaagtatdgtcat      | 777       |             |
| 33                    | 881   | GATTATTGTAANTTTCAGAGATAGAAATAAATAATTAATTAATTTTGAAAGGAAACTT    | 940       |             |
| 34                    | 778   | c   | 778       |             |
| 35                    | 941   | C   | 941       |             |

|  |   |  |
|--|---|--|
| RESULT   | 2 |  |
| US-08-418-071-5  |   |  |
| ; Sequence 5, Application US/08418071  |   |  |
| ; Patent NO. 5846705   |   |  |
| ; GENERAL INFORMATION:   |   |  |
| ; APPLICANT: Wu, Rey-Yuh   |   |  |
| ; APPLICANT: You, Li-Ru  |   |  |
| ; APPLICANT: Soong, Tai-Song   |   |  |
| ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  |   |  |
| ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING BANANA BUNCHY TOP VIRUS |   |  |
| ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  |   |  |
| ; NUMBER OF INVENTIONS: 18   |   |  |

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10020  
  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 5.0  
  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/418,071  
FILING DATE: 06-APR-1995  
CLASSIFICATION: 435  
  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: DCB-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1091 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
  
MOLECULE TYPE: subgenomic DNA  
DESCRIPTION: /desc "BBTV DNA II (clone 2)"  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
  
ORIGINAL SOURCE:  
ORGANISM: Banana Bunchy Top Virus (BBTV)  
  
US-08-418-071-5

| Query Match           | 18.3% | Score 142.6  | DB 2 | Length 1091                        |
|-----------------------|-------|--|------|------------------------------------|
| Best Local Similarity | 52.8% | Prod. No. 9.1e-38  |      |                                    |
| Matches               | 412   | Conservative   | 0    | Mismatches 349; Indels 20; Gaps 4; |
| QY                    | 1     | qtggggagcagggttgcaccttcaacttgcgcacgacacactccaaggattcaatccacttg   | 60   |                                    |
| Db                    | 173   | GTGCGCGCAGAGTCGCTCCGAGACCGGTGCGGAACGACTCCAGGAGATATCATCCCTG       | 232  |                                    |
| QY                    | 61    | aagacggttcggcgacgcaaggattgaagactgttcttgggaatgaacaggattcaactg     | 120  |                                    |
| Db                    | 233   | AGAANAATCTATTAAAGCTTTGGTAATTGAAGA---AGAGGTACTCTCGAAGCTCACTGG     | 289  |                                    |
| QY                    | 121   | gagcgcgaccctgggttccgcgcagacagaaatagagactactgttccgaaggaacggttgctt | 180  |                                    |
| Db                    | 290   | GAGAGGGCGAGGAGAACTGATGAACAGAATCGCAGATACTGTTTCGAAGGAAA-----CCC    | 344  |                                    |
| QY                    | 181   | ctcgaagcacgaggtcccgactgctctctggagtcacaaagccacgattgcccacgattt     | 240  |                                    |
| Db                    | 345   | CTTGAACTGGGTACTCCCGGTGTTCTCGTTTCGGAAGCGCAAGCTTCTCGATAGATTC       | 404  |                                    |
| QY                    | 241   | gctgaggaacactgatgaactcgcctctggaagaccacgcggcgataccgaagatgcgttgta  | 300  |                                    |
| Db                    | 405   | AGAGAGGCCCTGAGGAATTGAAGATGACGATCCATCCAAAGTATCGCATGCTTTGGCA       | 464  |                                    |
| QY                    | 301   | cacggagcttcggtgggaatggacaagatggccgctgaaatccggtcccatcttcocat      | 360  |                                    |
| Db                    | 465   | GTGGAATCAATTAAGATGCCAGAATTAAATTCGGAATGGGTTTCAGNACTAA-----        | 516  |                                    |
| QY                    | 361   | cacaattggcagcttgaagtgtctctgcgctgagagccagcgacgacgacgaatc          | 420  |                                    |
| Db                    | 517   | -AGAATGGCAAAATAAATTAATTCAACACATCGAAGGTGTTCTGTATGATCGAAGATATC     | 575  |                                    |
| QY                    | 421   | ctctggatgcgacgagacgaggaaggaacgggaagctccgtgttgccaaatattctcga      | 480  |                                    |
| Db                    | 576   | ATCTGGGTATACGGTCTCTCCGGAGGGCGAAGAAAGTCAACCTTCGCAAGATATCATCA      | 635  |                                    |

|    |     |  |     |
|----|-----|--|-----|
| Qy | 481 | ctcaagcccgactggttctacacatggttggaacagaaaagcagctattgtaccagttac   | 540 |
| Db | 636 | TTAAAACCTGGATGGGATATATCAACGGTGGAAAGACGTCGGATATCATGCATCATTA     | 695 |
| Qy | 541 | atcgaggagcccaaaacgaaatttaactcctcgatgtaccacagtgtaatttagagatttta | 600 |
| Db | 696 | ACGATGGATCTGTATAATCATTTGGATTATTGTATATCCCCAGAAGTCATTCAGATTATCTG | 755 |
| Qy | 601 | aattatgcccgtttagaatggtttaagaacagggcattcagttcggacaaatcagaacc-   | 659 |
| Db | 756 | AATTATGGCGTTATAGAACAAATTAAGAATACAGATTTTTAATAAATACAAAATACGAACCA | 815 |
| Qy | 660 | --cctagtgtatcttgggttcggaccattgcgatgcattcgtatttgcgaatgctcctgcct | 717 |
| Db | 816 | TGTGTGATTACAAAAGATGCACAAATGTCCATGTAAATGTATTGGCAAAATGTGTGGCT    | 875 |
| Qy | 718 | gattatttgaaaatcagcaggagcagaataaaactgtggaatatttaagtatgtgtcat    | 777 |
| Db | 876 | GATTATTGTAAATTTTCAGAAGATAGAATAAAAATAATTAATTCTGTGAGAAGGAAACTT   | 935 |
| Qy | 778 | c  | 778 |
| Db | 936 | C  | 936 |

```

RESULT      3
US-08-418-071-3
; Sequence 3, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

```

```

1  APPLICANT:  Marchall, Jerry S.
2  TITLE OF INVENTION:  NOVEL PLANT PROMOTERS AND USES
3  NUMBER OF SEQUENCES:  34
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Scully, Scott, Murphy & Presser
6  STREET:  400 Garden City plaza
7  CITY:  Garden City
8  STATE:  New York
9  COUNTRY:  U.S.A.
10 ZIP:  11530
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  Floppy disk
14 COMPUTER:  IBM PC compatible
15 OPERATING SYSTEM:  PC-DOS/MS-DOS
16 SOFTWARE:  PatentIn Release #1.0, Version #1.25
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER:  US/08/793,634B
19 FILING DATE:  June 9, 1997
20 CLASSIFICATION:  800
21
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  Digiglio, Frank S.
24 REGISTRATION NUMBER:  31,346
25 REFERENCE/DOCKET NUMBER:  10330
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE:  516-742-4343
28 TELEFAX:  516-742-4366
29
30 INFORMATION FOR SEQ ID NO:  2:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH:  1022 base pairs
33 TYPE:  nucleic acid
34 STRANDEDNESS:  single
35 TOPOLOGY:  linear
36
37 MOLECULE TYPE:  DNA (genomic)
38
39 IS-08-793-634B-2

```

|    |                           |   |
|----|---------------------------|---|
|    | Query Match               | 13.8%; Score 108; DB 4; Length 1022;                                      |
|    | Best Local Similarity     | 49.3%; Prod. No. 3e-26;   |
|    | Matches 378; Conservative | 0; Mismatches 370; Indels 18; Gaps 3;                                     |
| QY | 1                         | gttgaggacgaggttgcaccttcactccgaatggccaacgcaccctcacaagtattcatccacttg 60<br> |
| Dd | 178                       | GTTGTCGGCGACGAACAATGCAACTTACTGTGAGAGAACAACCCTCCAGGGAATTGTATCGTTTC 237<br> |
| QY | 61                        | aegacoggtcgccgactcaaggttaagaactgtttcttgggaatcacaggattcacctcg 120<br>      |
| Dd | 238                       | AAGAACAATAATTCGTTGGTGATTGAAGAAGAAAATTTGGTAATC---GAGCTCACTGG 294<br>       |
| QY | 121                       | gagccagcccgttgttcocgacgacacaaataagagactactgttcgaaggaacgggtgctt 180<br>    |
| Dd | 295                       | GAANAATGCCAGAGGCGAGCATCTCAGAATCCGATTATTGCTGTAAGAAGAACCCCTAATT 354<br>     |
| QY | 181                       | ctcagacgaggtccccactgccctctgagtcaaaaggccacgatlggcccaacgatit 240<br>        |
| Dd | 355                       | TCTGAATTGGGAATTCCCGGTCTAAGAGGGTTTCGAACRAGCGGAAGACGATGGAGATTAT 414<br>     |
| QY | 241                       | gctgaggaaacctgatgaactccgcttggaagaccaccgcgataccgaagatcgttgtta 300<br>      |
| Dd | 415                       | GAAGAGGATCCCGAGAAATGCAATTAAGGATCCAGATCTGCTCTTCGATGTGAAGCG 474<br>         |
| QY | 301                       | cacggagcttcggtggaaatggacaagatggccgcgtaaaaatccgttcccatcttcatat 360<br>     |
| Dd | 475                       | AAGAAATTTGAAGAGGAATATTGTTCTGTATTGATTTTCAGAAACCTCCGTCCA----- 528<br>       |
| QY | 361                       | cacaattggcagcttgaadtgtgtctcgatcggagagccagcgacgacatcgaccaatc 420<br>       |
| Dd | 529                       | -----TGGCAAATTGAGCTTCACGAGGATTTATPFGCGGAACCCAGATGATCGGAGTATC 582<br>      |
| QY | 421                       | ctctgaatgcgacgacgaggagagacgggaagtcggtgtttgccaatatctcgga 480<br>           |
| Dd | 583                       | ATCTGGGTCTATGGTTTCACAGCGAGGAGNAGNAAGACGAGCTTCGCCAAGGAATT--A 639<br>       |
| QY | 481                       | ctcaagcccagctggtttctacacatggtggtggaaccgaagaggacgtatgttacagtatc 540<br>    |

[illegible]

```

RESULT 5
US-08-793-634B-6
; Sequence 6, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.
; APPLICANT: Larkin, Philip J.
; APPLICANT: Taylor, William C.
; APPLICANT: Marchall, Jerry S.
; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08793.634B
; FILING DATE: June 9, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-793-634B-6

```

|                           |        |                    |           |              |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match               | 12.7%  | Score 99.2;        | DB 4;     | Length 1017; |
| Best Local Similarity     | 54.1%; | Pred. No. 2.5e-23; |           |              |
| Matches 225; Conservative | 0;     | Mismatches 188;    | Indels 3; | Gaps 1;      |

[illegible]

```

RESULT      7
US-08-973-068-28
; Sequence 28, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; CURRENT FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; EARLIER FILING DATE: 1996-05-31
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 982
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-28

```

[illegible]

[illegible]

|     | Query Match  | 7.6%            | Score 59.6;         | DB 1;     | Length 1110; |
|-----|--|-----------------|---------------------|-----------|--------------|
|     | Best Local Similarity  | 51.3%;          | Pred. No. 4,2e-104; |           |              |
|     | Matches 193;   | Conservative 0; | Mismatches 174;     | Indels 9; | Gaps 2;      |
| 378 | agtgcgtctgcgtcggagagcaggcaggcggatccacaaacctcctcgatcgcggacg       | 437             |                     |           |              |
|     |  |                 |                     |           |              |
| 584 | AGAGGTGGAGAATATCATGGCGCAGCCGATGCTATCGAGAAATAAATTTGGGTCTATGGACC   | 643             |                     |           |              |
| 438 | agacggagagacggcgggaagtcgctgtgttcgcaaatatctcggactcaagcccgacttggtt | 497             |                     |           |              |
|     |  |                 |                     |           |              |
| 644 | AAATGGAGGAGGAGGAAGACACCGTATGCAAAACATCTTAATGAAGACGAGAAATCGGTT     | 703             |                     |           |              |
| 498 | ctacacatgtgtggaaaccaggaaggcgtattgtaccagtcacatcagggaccccaaacg     | 557             |                     |           |              |
|     |  |                 |                     |           |              |
| 704 | TATATCTCCAGGAGAAATCATTTGGATATATCTAGACTGTATTAATATACGAGGATAT---    | 760             |                     |           |              |





|    |     |   |     |
|----|-----|---|-----|
| Qy | 438 | aqacgaggaacgcggaagtccgtgtttgccaaatatctcgactcaatcgaagccccactggtt | 497 |
| Dd | 671 | AAATGGAGGACGAAGAACAACCGTATGCAAAACATCTAATGAAGACGAGAAATCCGTT      | 730 |
| Qy | 498 | ctacacatgtgttggaaaccagaagagcatttgtaccagtcatactcaggaccccacaacg   | 557 |
| Dd | 731 | TTATTCTCAGGAGGAAAATCATTTGGATATATGTAGATGTATAATTACGAGGATAT---     | 787 |
| Qy | 558 | aaatttaactctcgatgtaccagggtgaatttagagtatattaattatccctgttaga      | 617 |
| Dd | 788 | ---TGTTATTTTCATATCCAAGATSCAAAGAGGATTATTAAATTATGGGTTATTAGA       | 844 |
| Qy | 618 | atgtgttaagAACGGGCatttcagtttcgacaaaatacgaaccccttgattcatcttgggt   | 677 |
| Dd | 845 | GGAATTTAAGTAATGGAATANTTCAGACGGGAATAATGACCCTGTTTGAAGATAG--T      | 901 |
| Qy | 678 | cgacctatgcgatgtactcgtattttgccaatgtccctgcctgattatttgaataatcagcag | 737 |
| Dd | 902 | AGAATATGCGAAGTCATGTGTGGCTAAC'TTCCTCCGAAGGAAGAAATCTTTTCTGA       | 961 |
| Qy | 738 | ggacagaataaaactg  | 753 |
| Dd | 962 | AGATCGAATAAAGTTG  | 977 |

```

RESULT 12
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

```

Query Match 7.5%; Score 58.6; DB 1: Length 1103;  
Best Local Similarity 52.5%; Pred. No. 9.1e-10;  
Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;

QV 410 atcgcacactctcttgatcgagcagacgaggagacgggaagtccdtgtttcca 469

616 ATCGAAGGATTATTTGGGTCTACGGCCCAAAATGGAGGCGAAAGGAAACAAACGCTTTTGCAA 675  
 470 atatctcggactcaagcccgactggtttctacacatggtggtgggaaccagaagaagacgtat 529  
 676 AACATTTAATGAAGACTAAGAATGCGTTTTATTGCGCAGGAGGAATAATCATTTGGATATAT 735  
 530 tgtaccagttacatcgagagcccaaaacgaatttaatcctcgatgtacccagggtgtaatt 589  
 736 GTAGATTGTATAATTATGAGGATATA-----GTTATATTGTATTTCCAGATGTCNAAG 789  
 590 tagagtatttaaatatgctcctgtagaatggttaagaacacagggcattcagttcggaca 649  
 790 AGGAATTTTAACTATGCGCTTATTAGAGAAATTTAAAAATGCAATTTTCAAAAGCGGA 849  
 650 aataagaaccccttagttatcttgggttcgacatgfcgatgtactcgtatttgccaatg 709  
 850 AATATGAACCCGTTTGTAAATTTG---TAGAATATGTGGAATCATTGTAATGCTAACT 906  
 710 tctcgctgattatttgaataatcagcagggcagataaaact 752  
 907 TCTTCCGAGCAAGGAATCTTTTCTCAAGATCGAATAAAGCT 949

```

RESULT 13
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-18

```

Query Match 7.5%; Score 58.6; DB 1; Length 1105;  
Best Local Similarity 52.5%; Pred. No. 9, 1e-10;  
Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;  
QY 410 atcgacaatctcttgatcgcgagacgaggagacgggaagtccgtgtttgcca 469





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 09:56:55 ; Search time 180.6 Seconds  
(without alignments)  
1353.271 Million cell updates/sec

Title: US-09-462-955-1  
Perfect score: 1291  
Sequence: 1 ccgcacaaacctctgctaag.....ggacgctgagtgatctgg 1291

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description       |
|------------|-------|---------------|--------|-------|-------------------|
| 1          | 179.2 | 13.9          | 1096   | 2     | US-08-418-071-4   |
| 2          | 163.4 | 12.7          | 1106   | 2     | US-08-418-071-3   |
| 3          | 159.6 | 12.4          | 1091   | 2     | US-08-418-071-5   |
| 4          | 112.4 | 8.7           | 1022   | 4     | US-08-793-634B-2  |
| 5          | 99.2  | 7.7           | 1017   | 4     | US-08-793-634B-6  |
| 6          | 62.8  | 4.9           | 1110   | 1     | US-08-202-186-14  |
| 7          | 59.6  | 4.6           | 982    | 3     | US-08-973-068-28  |
| 8          | 59.6  | 4.6           | 1110   | 1     | US-08-202-186-11  |
| 9          | 59.6  | 4.6           | 1111   | 1     | US-08-202-186-9   |
| 10         | 59.6  | 4.6           | 1111   | 1     | US-08-202-186-12  |
| 11         | 59.6  | 4.6           | 1111   | 1     | US-08-202-186-24  |
| 12         | 58.6  | 4.5           | 1103   | 1     | US-08-202-186-16  |
| 13         | 58.6  | 4.5           | 1105   | 1     | US-08-202-186-18  |
| 14         | 58    | 4.5           | 1109   | 1     | US-08-202-186-13  |
| 15         | 58    | 4.5           | 1111   | 1     | US-08-202-186-15  |
| 16         | 57    | 4.4           | 1104   | 1     | US-08-202-186-17  |
| 17         | 56.4  | 4.4           | 1111   | 1     | US-08-202-186-10  |
| 18         | 46.4  | 3.6           | 287    | 2     | US-08-418-071-1   |
| 19         | 46.4  | 3.6           | 300    | 2     | US-08-418-071-2   |
| 20         | 37.2  | 2.9           | 7218   | 1     | US-08-232-463-14  |
| 21         | 31.8  | 2.5           | 498    | 1     | US-07-781-254A-14 |
| 22         | 31.8  | 2.5           | 1338   | 1     | US-07-781-254A-6  |
| 23         | 31.8  | 2.5           | 1338   | 1     | US-07-781-254A-7  |
| 24         | 31.8  | 2.5           | 1863   | 1     | US-07-781-254A-4  |
| 25         | 31.8  | 2.5           | 1863   | 1     | US-07-781-254A-5  |
| 26         | 31.2  | 2.4           | 1768   | 4     | US-09-347-594-3   |
| 27         | 31.2  | 2.4           | 1768   | 4     | US-09-347-594-4   |

|    |      |     |      |   |                  |                   |
|----|------|-----|------|---|------------------|-------------------|
| 28 | 31   | 2.4 | 1212 | 3 | US-09-046-578-5  | Sequence 5, Appli |
| 29 | 30.8 | 2.4 | 265  | 3 | US-08-289-222E-5 | Sequence 5, Appli |
| 30 | 30.8 | 2.4 | 265  | 4 | US-09-054-526B-5 | Sequence 5, Appli |
| 31 | 30.8 | 2.4 | 1767 | 4 | US-09-347-594-1  | Sequence 1, Appli |
| 32 | 30.8 | 2.4 | 1767 | 4 | US-09-347-594-2  | Sequence 2, Appli |
| 33 | 30.8 | 2.4 | 2272 | 1 | US-08-482-577B-1 | Sequence 1, Appli |
| 34 | 30.8 | 2.4 | 2272 | 3 | US-08-289-222E-2 | Sequence 2, Appli |
| 35 | 30.8 | 2.4 | 2272 | 4 | US-09-218-176-1  | Sequence 1, Appli |
| 36 | 30.8 | 2.4 | 2272 | 4 | US-09-054-526B-2 | Sequence 2, Appli |
| 37 | 30.4 | 2.4 | 1212 | 3 | US-09-046-578-1  | Sequence 1, Appli |
| 38 | 30   | 2.3 | 1437 | 3 | US-08-724-814-15 | Sequence 15, Appl |
| 39 | 30   | 2.3 | 3182 | 1 | US-08-188-582-1  | Sequence 12, Appl |
| 40 | 30   | 2.3 | 3182 | 1 | US-08-646-715-12 | Sequence 12, Appl |
| 41 | 29.8 | 2.3 | 951  | 1 | US-08-671-525B-1 | Sequence 1, Appli |
| 42 | 29.8 | 2.3 | 951  | 1 | US-08-672-109B-1 | Sequence 1, Appli |
| 43 | 29.8 | 2.3 | 951  | 1 | US-08-842-045-1  | Sequence 1, Appli |
| 44 | 29.8 | 2.3 | 951  | 2 | US-08-842-238-1  | Sequence 1, Appli |
| 45 | 29.8 | 2.3 | 951  | 3 | US-08-629-335B-1 | Sequence 1, Appli |

ALIGNMENTS

RESULT 1  
US-08-418-071-4  
; Sequence 4, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh  
; APPLICANT: Soong, Tai-Seng  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/418.071  
; FILING DATE: 06-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: DCB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1096 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; MOLECULE TYPE: genomic DNA  
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Banana Bunchy Top Virus (BBTV)  
US-08-418-071-4

Thu Oct 25 13:07:51 2001

```

Query Match      13.9%; Score 179.2; DB 2; Length 1096;
Best Local Similarity 52.4%; Pred. No. 2e-50;
Matches 499; Conservative 0; Mismatches 433; Indels 20; Gaps 4;

QY 40 agccgccccgtagtattaccocccgcccccccaactctgtacaccccgcttgcta 99
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 AGGAGGAGCGGCTAGTATTACCGCTCTCTCGCACCTTCCTCGCACCTGACGTC-- 64

QY 100 taatgggttcctccattccgctggtgtctcactctgaactcagacagaggaagag 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 ---ATCATTTATCTCTTTTAAATGGTGTCTCACTCTCAATTTATCTCTCGCAGCGGAG 121

QY 160 gcggcgaaatgttgtaagcgcatcgatgctcccttaacttgggtctatgtatgttggggac 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 CGAGAAGACTTCTCGCTCTTCTGAAGGAGGAGGATGTTCACTACTCTGTCTGCTGGCGAC 181

QY 220 gaggttgacacttcacagcagacacaccccaagattaccacactcactgaagaccggt 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 GAAGTCCGCTCCCGCCACCGCCAGAGACCTCCAGGGATATCTATCCCTGAANAATCA 241

QY 280 cggcgaactcaagattgaagactgttcttgggaatgacagagattcactcgtggagccgacc 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 ATTCCGCTCGCGGATTGAAAAGAAGTATGG---CTCCGCTCACTGGGAGATTGCG 298

QY 340 cgtggttcgacagacagaaatgagactactgttcgaaggaacgggtgcttcagagcac 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 AAAGGAAGTCAACAGAGATCGCATACTGTTCGAAGGAACCCCTAGTCTTGAACCTG 358

QY 400 gaggtccgactcgtcctgagtcgaagggccacgattggcccaacgatttctgctgagaa 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 GGTACTCCGGTGTCTGCTGGTTCGAGAGACGCAAGCTTCTGATAGATTACAGAGAGAGC 418

QY 460 cctgtagaacctccgctggaagaccagcgcgatgatacgaagatgctgtgacacggagct 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 CTTGAGGAATTGAAGATGGACGATCTCAAGTATCGCAGATGCTTGGCAGTGGAAATCA 478

QY 520 tcggtggaatggacaagatgggcccgtgaaatccgttcccttccatcacatctg 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 ATTAAAGATGCCAGAAATTAATTCGGAATGGGTTCAGCACTA-----AAAGAATGG 529

QY 580 cagctgaagtgtgtctgcgactcgagagccagcgagcagatcgacacatcctctgata 639
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 CAAATAAATAATTAATCAACACATCGAAGGTGTTCTCGATGATCGAAGATATCATCTGGGTA 589

QY 640 tgcgagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 699
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 TACGGTCCCAACAGCGGCGGAGGAAAGTCAACCTTCGCAAGATATCTATCAITFAAACCC 649

QY 700 gactggttctacacatgtgtggaaccagaaagacgagctattgtaccagttacatcgaggac 759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 GGATGGGATATATCAACGGTGGAAAGACGCTCGGATATGATGATGATGATGATGATGATGAT 709

QY 760 ccaaacagaaatttaactcctcgatgataccagcagcagcagcagcagcagcagcagcagc 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 710 CCTGATAATCATTTGGATTATGATATCCAGAAAGTCACTTCAAGATATCTGAATTTATGGC 769

QY 820 ctgttaaatgtttaagaacagcgactcagttcggacacaaatcacgaacc---ccttagt 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 770 GTTATAGAACAAATTAAGATAGATGATTTTAAATAATACAAATAGCAACCATGTGTGATT 829

QY 877 tatctgggttcgaccatgtgcatgtactcgtatttgcgaatgtcctcgtcgtatttg 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 AGAAAAGATGCAAAATATGCTCATGTAATTTATGTTATGCAAAATGTGTCCTGATTTATGT 889

QY 937 aaaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 AAAATTTCCAGAGATAGAAATAAATAATTAATTTCTTGAGAAAGCAACTTC 941

```

RESULT 2

US-08-418-071-3

; Sequence 3, Application US/08418071

; Patent No. 5846705

```

GENERAL INFORMATION:
APPLICANT: Wu, Rey-Yuh
APPLICANT: You, Li-Ru
APPLICANT: Soong, Tai-Seng
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,071
FILING DATE: 06-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: DCB-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1106 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: genomic DNA
DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-3

```

```

Query Match      12.7%; Score 163.4; DB 2; Length 1106;
Best Local Similarity 50.6%; Pred. No. 4.3e-45;
Matches 454; Conservative 0; Mismatches 431; Indels 12; Gaps 2;

QY 83 ctaaccccgcttggtataaataagggttctcattccgctgctgctcactcactgaact 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42 CTACCTCTGAACGCTGGAGATGTCCTCCTCTTAAAGTGGTGGTCTCACTCTGATTT 101

QY 143 acgagacagagaaagagcgcgcaatgttgaaggcgcatcgatcccttaacttggct 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102 ACTCTCCGCGCACAGAGAGAAACTTCTCTCTCTTCTGAAGGAGGAGGATGTTCACT 161

QY 203 atgtatagtgtgggacaggttgccactcactgacacacacacacacacacacacacac 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 ACGCTGTCTGCGGACGAAAGTCTGCTCGGCGCACGAGGAGGAGGAGGAGGAGGAGGAGG 221

QY 263 tccactgaagaccggtcgcgactgcaaggttgaagactgttcttgggaatgacagga 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 TATCCCTGAAAAGAGAAATCCCGCTCGGCGGATTTGAAGAGAGATATGG---TTCCCGCTG 278

QY 323 ttcacactggagccgacccctggttccgcagacacacacacacacacacacacacacac 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 CTCACCTGGGAGATGTCAGAGGAAACGAGAAATTCGAAGTACTGTTTCAAAAGAAA 338

QY 383 ggggtcttcagacacgaggtcccgactcgtccctggagtcacaaagggccacgattggccc 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 CCCTAATCTCGAATTAGGGTTTCTGTTTAAATGGTCTTAATAAAGAGAAATATCGG 398

```



|     |    |   |     |
|-----|----|---|-----|
| 204 | yy | TTGACAGAAACACCTCCAGGATTTGTATCGTTCAAGAACAAAATTCGTCTTGGTGGATT       | 263 |
| 297 | yy | gaagactgtctctgggaatgacaggaattcacctgagccgcgacccgtggtcccgacgaaca    | 356 |
| 264 | bb | GAAGAGAAATTTGGTAATC---CAGCTCACTGGGAAATTCGGAGAGCAGCGATTCTCA        | 320 |
| 357 | yy | gaatagagactactgttccgaaggaacgggtgctcttcgacgacgcgagtcgccgactcgctc   | 416 |
| 321 | bb | GAATCGCGATTATTCCTGTAAAGAAACCCATAATTCGTGAAATGGGATTCGGTCATGAA       | 380 |
| 417 | yy | tgagtcacaaagcccaagcattggcccacgatttctgtaggaacctgatgaactccgcct      | 476 |
| 381 | bb | GGTTTCGACACAGCGGAGACGATGGAGATTTATGAAGAGGATCCCGAAGAAATGCAATT       | 440 |
| 477 | yy | ggaagaccagcggtataccgaagaatcgctgtgtacacgagcgtccggtggaatgacaag      | 536 |
| 441 | bb | GAAGGATCCAGATACTGCTCTTCGATGTAAAGCGAAGAAATTCGAAAGAGGAATATTGGTC     | 500 |
| 537 | yy | atgggcgcgtgaaatccgctcccatctccatattccataaccaaattgacagcttgaagtcgtgc | 596 |
| 501 | bb | CTGTTATGATTTTCAGAACTCCGT-----CCATGGCAAAATTCAGCTTCACGA             | 548 |
| 597 | yy | tcgcatcgagagccagcggagcgtgcacaaatcctctgatatcggcgacgagacggagg       | 656 |
| 549 | bb | GGATTTAATGGCGAACCCAGATGATCGGAGTATCATCTGGCTATATGGTTTCAGACGGAGG     | 608 |
| 657 | yy | agacgggaagtcocggtgtttgccaataatctcggactcaacgcccgactgtttctcacatg    | 716 |
| 609 | bb | AGAAGGAAAGACGAGCGTTCGCGAAGGAATT---AATCAGGTATGGATGGTTTATACAGC      | 665 |
| 717 | yy | tggtgaacacagaagggcgtattgtaccagtcacatcgagaccacccaaacgaatttaatt     | 776 |
| 666 | bb | CGGAGGGAAGACCCAGACGTTATATATATGTATGCTCAAGACCCAGAGGGAATATTCG        | 725 |
| 777 | yy | octogatgtaccaggtgaatttagagattttaaatatgcctcttagaatgtgttaa          | 836 |
| 726 | bb | GTTTGATGTTCCACGCTGTTCTCGGAGATGATGAACATATACGGCATGTGAGATGTGAA       | 785 |
| 837 | yy | gaacagggcattcagttcgggacaaatacgaaccccttagttatcttgggttcgaccatgt     | 896 |
| 786 | bb | GAACACAGAGTTTTTGC AAGTACAAAATATATAGGCCTGTAGTCCTTGTATATTAGGAAGT    | 845 |
| 897 | yy | gcattactcgtatttggcgaalgtcctgcctgattatttgaataatcagcagggacagaat     | 956 |
| 846 | bb | TCATTTAAATTGTTTGGCCAACTGGCACCTGACCCCGCCATAACTGACGACAGACT          | 905 |
| 957 | yy | aaaactgtggaattattaa   | 976 |
| 906 | bb | TGTAATTATCAATTGTTGAA  | 925 |



RESULT 7  
US-08-973-068-28  
; Sequence 28, Application US/08973068  
; Patent No. 6127604  
; GENERAL INFORMATION:  
; APPLICANT: Dale, James Langham  
; APPLICANT: Harding, Robert Maxwell  
; APPLICANT: Dugdale, Benjamin  
; APPLICANT: Beetham, Peter Ronald  
; APPLICANT: Hafner, Gregory John  
; APPLICANT: Becker, Douglas Kenneth  
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS  
; FILE REFERENCE: 09457/002001

7

US-08-973-068-28

: Sequence 28, Application US/08973068

; Patent No. 6127604

; GENERAL INFORMATION:

APPLICANT: Dale, James Langham

APPLICANT: Harding, Robert Maxwell

APPLICANT: Dugdale, Benjamin

APPLICANT: Beetham, Peter R

APPLICANT: Hatner, Gregory John

APPLICANT: Becker, Douglas Kenneth

FILE OF INVENTION: INTERGEN;  
FILE REFERENCE: 00657/003001

NAME: JEFFERY, Donald D.  
REGISTRATION NUMBER: 19,980  
REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1110 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
US-08-202-186-11

Query Match 4.6%; Score 59.6; DB 3; Length 982;  
Best Local Similarity 51.3%; Pred. NO. 4.5e-10;  
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 588 agtgcgtctgcgacgagagccagcgagcgcacacatctctggtatgcggagc 647  
Db 53 agagtggaataatcatggcgccatctgcgaggaataatttgggtctatggccc 112  
QY 648 agacgagagagcggaagtcggtttgccaataatctcggaactcaagcccgactggt 707  
Db 113 aaatgagagaggaagacacacgtatgcaaaacatctaatgaagacgagaaaatgcgt 172  
QY 708 ctacacatggtggaaccagagagacgtattgtaccagatcacatcgaggagcccaaaacg 767  
Db 173 ttattccagaggaaaatcatcttgatattgtatgactgtataattacagagatat--- 229  
QY 768 aaatttaactcctgactgaccaggtgttaatttagagatttaaaattatgcctgttaga 827  
Db 230 ---tgattattgatttccagatgcaaaagagattttaaattatgggtatttaga 286  
QY 828 atgtgtaagacagggcattcagttcggacaaataacgaaccccttagtaccttgggtt 887  
Db 287 ggaatttaagaatggaataattcaagcgaggaaatgaacccgttttgaagatag---t 343  
QY 888 cgaccatgctgactgactgtatttccaaatctcgcctgattatttgaaatcagcag 947  
Db 344 agaataatgcaagtcattgtaattggtcaactcctcccgagggaaggaatttttctga 403  
QY 948 ggacagaataaaaactg 963  
Db 404 agatcgaataaaattg 419

RESULT 8  
US-08-202-186-11  
; Sequence 11, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

QY 588 agtgcgtctgcgacgagagccagcgagcgcacacatctctggtatgcggagc 647  
Db 584 AGAGTGGAGAAATCATGCGCGAGCCATGTCATCGAGAAATAATTTGGGTCTATGGACC 643  
QY 648 agacgagagagcggaagtcggtttgccaataatctcggaactcaagcccgactggtt 707  
Db 644 AATGGAGGAGAGAGAAAGACAACTGATGCAAAACATCTAATGAAGACGAGAAATCGGTT 703  
QY 708 ctacacatggtggaaccagagagacgtattgtaccagatcacatcgaggagcccaaaacg 767  
Db 704 TTATTCCTCCAGGAGGAAATCATTGGATATATGTAGACTGTATAATTCAGGATAT--- 760  
QY 768 aaatttaactcctgactgaccaggtgttaatttagagattttaaattatgcctgttaga 827  
Db 761 ---TGTATATTTCATATCCAGATGCAAAAGAGATTATTAAATTTATGGTTATTAGA 817  
QY 828 atgtgtaagacagggcattcagttcggacaaataacgaaccccttagtaccttgggtt 887  
Db 818 GGAATTTAAGAATGGAATAATTCAAAGCGGAAATATGAACCCGTTTGAAGATAG---T 874  
QY 888 cgaccatgctgactgactgtatttccaaatctcgcctgattatttgaaatcagcag 947  
Db 875 AGAATATGCGAAGTCATTGTAATGGCTACTTCTTCGGAAGGAGGAATCTTTCTCTGA 934  
QY 948 ggacagaataaaaactg 963  
Db 935 AGATCGAATAAAGTTG 950

RESULT 9  
US-08-202-186-9  
; Sequence 9, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186



RESULT 12  
US-08-202-186-16  
; Sequence 16, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA

```

RESULT 13
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
;

```

COMPUTER READABLE FORM:

STREET: 3000 K Street, N.W.

RESULT 15  
US-08-202-186-15  
Sequence 15, Application US/08202186  
Patent No. 5756708  
GENERAL INFORMATION:  
APPLICANT: KARAN, Mirko  
APPLICANT: BURNS, Thomas M.  
APPLICANT: DALE, James L.  
APPLICANT: HARDING, Robert M.  
TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington, D.C.

Search completed: October 24, 2001, 10:00:12  
Job time: 197 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:16 : Search time 393.52 Seconds  
(without alignments)  
1246.166 Million cell updates/sec

Title: US-09-462-955-1-copy\_211\_991

Perfect score: 781

Sequence: 1 gttgggacaggtgtgacc.....ttaaagtatgtgtcatctaa 781

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 162.6 | 20.8        | 1096   | 20 | AAV71833    |
| 2          | 162.6 | 20.8        | 1096   | 21 | AAA38950    |
| 3          | 145.4 | 18.6        | 886    | 20 | AAA34687    |
| 4          | 142.6 | 18.3        | 1091   | 20 | AAV71834    |
| 5          | 142.6 | 18.3        | 1091   | 21 | AAA38947    |
| 6          | 137.8 | 17.6        | 1106   | 20 | AAV71832    |
| 7          | 136.2 | 17.4        | 1106   | 21 | AAA38946    |
| 8          | 108   | 13.8        | 1022   | 17 | AAT13161    |
| 9          | 99.2  | 12.7        | 1017   | 17 | AAT13165    |
| 10         | 80.4  | 10.3        | 593    | 20 | AAA34686    |
| 11         | 62.8  | 8.0         | 1110   | 19 | AAV24089    |

|    |      |     |      |    |          |                    |
|----|------|-----|------|----|----------|--------------------|
| 12 | 59.6 | 7.6 | 982  | 18 | AAV49405 | Banana bunchy top  |
| 13 | 59.6 | 7.6 | 1110 | 19 | AAV24086 | Banana bunchy top  |
| 14 | 59.6 | 7.6 | 1111 | 19 | AAV24077 | Banana bunchy top  |
| 15 | 59.6 | 7.6 | 1111 | 19 | AAV24084 | Banana bunchy top  |
| 16 | 59.6 | 7.6 | 1111 | 19 | AAV24087 | Banana bunchy top  |
| 17 | 58.6 | 7.5 | 1103 | 19 | AAV24091 | Banana bunchy top  |
| 18 | 58.6 | 7.5 | 1105 | 19 | AAV24093 | Banana bunchy top  |
| 19 | 58   | 7.4 | 1109 | 19 | AAV24088 | Banana bunchy top  |
| 20 | 58   | 7.4 | 1111 | 19 | AAV24090 | Banana bunchy top  |
| 21 | 57   | 7.3 | 1104 | 19 | AAV24092 | Banana bunchy top  |
| 22 | 56.4 | 7.2 | 1111 | 19 | AAV24085 | Banana bunchy top  |
| 23 | 46.4 | 5.9 | 287  | 20 | AAV71830 | Subgenomic fragmen |
| 24 | 46.4 | 5.9 | 287  | 21 | AAA38948 | Banana bunchy top  |
| 25 | 46.4 | 5.9 | 300  | 20 | AAV71831 | Subgenomic fragmen |
| 26 | 46.4 | 5.9 | 300  | 21 | AAA38949 | Banana bunchy top  |
| 27 | 45.8 | 5.9 | 936  | 22 | AAF58252 | Banana bunchy top  |
| 28 | 45.8 | 5.9 | 936  | 22 | AAF58254 | Oligonucleotide D1 |
| 29 | 45.8 | 5.9 | 936  | 22 | AAF58257 | Oligonucleotide D1 |
| 30 | 45.8 | 5.9 | 936  | 22 | AAF58259 | Oligonucleotide D1 |
| 31 | 45.8 | 5.9 | 936  | 22 | AAF58262 | Oligonucleotide D1 |
| 32 | 45.8 | 5.9 | 936  | 22 | AAF58265 | Oligonucleotide D1 |
| 33 | 42.4 | 5.4 | 936  | 22 | AAF58252 | Oligonucleotide D1 |
| 34 | 42.4 | 5.4 | 936  | 22 | AAF58254 | Oligonucleotide D1 |
| 35 | 42.4 | 5.4 | 936  | 22 | AAF58257 | Oligonucleotide D1 |
| 36 | 42.4 | 5.4 | 936  | 22 | AAF58259 | Oligonucleotide D1 |
| 37 | 42.4 | 5.4 | 936  | 22 | AAF58262 | Oligonucleotide D1 |
| 38 | 42.4 | 5.4 | 938  | 22 | AAF58255 | Oligonucleotide D1 |
| 39 | 37   | 4.7 | 244  | 22 | AAF58238 | Oligonucleotide D1 |
| 40 | 32.6 | 4.4 | 244  | 22 | AAF58238 | Oligonucleotide D1 |
| 41 | 32.6 | 4.2 | 5053 | 19 | AAV22816 | Nucleotide sequenc |
| 42 | 32.6 | 4.2 | 5055 | 19 | AAV17236 | DNA from a region  |
| 43 | 32.4 | 4.1 | 1149 | 21 | AAA51838 | Human melanocortin |
| 44 | 32.4 | 4.1 | 1540 | 21 | AAA51837 | Human melanocortin |
| 45 | 31.8 | 4.1 | 1863 | 12 | AAQ14271 | D-3 dopaminergic r |

## ALIGNMENTS

RESULT 1  
AAV71833  
ID AAV71833 standard; DNA; 1096 BP.  
XX  
AC AAV71833;  
XX  
DT 10-FEB-1999 (first entry)  
XX  
DE BBTV DNA II clone (2-17) nucleotide sequence.  
XX  
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
XX  
OS Banana bunchy top virus.  
XX  
FH Key Location/Qualifiers  
FT TATA\_signal 1..7  
FT /\*tag= a  
FT stem\_loop 8..38  
FT /\*tag= b  
FT CDS 70..928  
FT /\*tag= c  
FT /product= "ORF-V2 product"  
FT /transl\_except= (pos:215..217, aa:Gly)  
FT polyA\_signal 533..538  
FT /\*tag= d  
FT polyA\_signal 799..804  
FT /\*tag= e  
FT polyA\_signal 907..912  
FT /\*tag= f  
FT polyA\_signal 1030..1035  
FT /\*tag= ge  
XX  
PN US5846705-A.







Db 595 gttgctgattattgtaaaatttcagaagatagataataataattattgtga 649

RESULT 4

AAV71834

ID AAV71834 standard; DNA; 1091 BP.

XX

AC AAV71834;

XX

DT 10-FEB-1999 (first entry)

XX

DE BBTV DNA II clone (2) nucleotide sequence.

XX

KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana; Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.

XX

OS Banana bunchy top virus.

XX

PN US5846705-A.

XX

PD 08-DEC-1998.

XX

PF 06-APR-1995; 95US-0418071.

XX

PR 06-APR-1995; 95US-0418071.

XX

PA (BIOT-) DEV CENT BIOTECHNOLOGY.

XX

PI Soong T, Wu R, You L;

XX

PT WPI; 1999-059037/05.

XX

PS Nucleic acids having banana bunchy top virus component sequences -

XX

PT used to design primers for use in polymerase chain reaction

XX

PT detection of the virus

XX

PS Disclosure; Fig 12A-C; 27pp; English.

XX

CC This represents the nucleotide sequence of a banana bunchy top virus (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention provides nucleic acid sequences associated with BBTV that can be used in a PCR technique for detecting BBTV. The nucleic acid sequences (AAV71830 to AAV71833) are used as the basis for the construction of PCR primers, to detect BBTV infection. The PCR technique is used for detecting BBTV in plant tissues (preferably banana, especially Musa species). The virus, one of the most important banana species viruses, causes phloem damage and is transmitted by aphids. PCR detection gives accurate, reliable and specific determination of absence or presence of the virus.

XX

SQ Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other;

Query Match 18.3%; Score 142.6; DB 20; Length 1091;  
Best Local Similarity 52.8%; Pred. No. 4.5e-38;  
Matches 412; Conservative 0; Mismatches 349; Indels 20; Gaps 4;

QY 1 gttgggacaggttgacacttcaactgcccacacgacactccaaggattccacttg 60  
Db 173 gtccggcaggaagtgcctccgagcaccgctcggaacacactccaggatattccctg 232  
QY 61 aagaccggtcggcagctgcaagattgaagactgttcttgggaatgacaggattccactg 120  
Db 233 aagaactcttaagcttggtggtgaaga---agaggactcttcgaaggtcactgg 289  
QY 121 gagccgcccgtggtccgacgaacagaatagagactactgttcgaagaacgggtgctt 180  
Db 290 gagagggcgaggaactgatgacagatcgcagatcactgttcgaaggaaa-----ccc 344  
QY 181 ctgagcacagagtcctccgactcgtctggagtcataaaggccacgattggcccaacgatt 240  
Db 345 cttgaactgggtactccggtggttcctggttcgaagaagcgcaagcttcctcgatattc 404

QY 241 gctgaggaacctgatgaactccgctcggaagaccgagcggtgataccgaagatgcgttgta 300  
Db 405 agagagagccctgaggaattgaagatggacgacatccatccaagtatcgagatgcttgga 464  
QY 301 cagcgagcttcggtggaatggacaagatggcgctgaaaaatccctccatttccattat 360  
Db 465 gtggaatcaattaaagatgccagaattaatccgaatgggttcacgaactaa----- 516  
QY 361 cacaattggcagctgaagtgtctgcgacgcgagagccagcgagatcgcaaatc 420  
Db 517 -aagaatggcaaaataaatttaacacacatcgaaaggtgttccctgatgatcgagatc 575  
QY 421 ctctgatatgcgacgacgagagagcggaagcgggaaagcctggtttgccaataatctcga 480  
Db 576 atctggttatacgttctccgagagcggaagaaagtcacacctcgcaagatattatca 635  
QY 481 ctcaagcccgactggtttacacatgtgtggaacacagaaaggaagcgtattgtaccagta 540  
Db 636 ttaaacctggatgggatatatacaacggtggaaagcgtcgatatgatgcacatcata 695  
QY 541 atcgagagaccccaaaacgaatttaactcctcgatgtaccaggtgtaatttagagtatta 600  
Db 696 acgatgactcctgataatcattggtattgtattgatatcccgagaagtcattcagattatcg 755  
QY 601 aattatgcccctgttagaattgttaagaacagggcattccagttcggacacaaatcacgaac- 659  
Db 756 aattatgcttatagaacaaattaaagatatagatttttaataataacaaatcacgaacca 815  
QY 660 --ccttagttattcttggttcgaccatgtgcattgctactgtatttgcgaatgcctgcct 717  
Db 816 tgtgtattagaaaagatggacaaaatgtccatgttaattgttggcaaatgctgtgctt 875  
QY 718 gattattgaaatccagcagggacagaataaaaactgtggaatatttaaagatgtgtcat 777  
Db 876 gattattgaaatccagcagggacagaataaaaactgtggaatatttaaagatgtgtcat 935  
QY 778 c 778  
Db 936 c 936

RESULT 5

AAA38947

ID AAA38947 standard; DNA; 1091 BP.

XX

AC AAA38947;

XX

DT 23-AUG-2000 (first entry)

XX

DE Banana bunchy top virus related nucleotide sequence #2.

XX

KW Banana bunchy top virus; BBTV; detection; ds.

XX

OS Banana bunchy top virus.

XX

PN TW360710-A.

XX

PD 11-JUN-1999.

XX

PF 30-JUN-1994; 94TW-0106105.

XX

PR 30-JUN-1994; 94TW-0106105.

XX

PA (BIOT-) DEV CENT BIOTECHNOLOGY.

XX

PI Wu R, You L, Song T;

XX

DR WPI; 2000-316145/27.

XX

PT Two circular single-stranded DNAs associated with banana bunchy top virus and detection of the virus.

XX

PS Claim 2; Page 1; 7pp; Chinese.



```

Db 230 aaaaagagaatccgctcgcgagattgaagaagatg---ttccggtgctcactgg 286
Qy 121 gagcgcacccgtggttcgcagacacagaatagacactactgttcgaaggaacgggtgctt 180
Pt 287 gagattgcagaggaaacgacgagagaaattcgaagtactgttcgaaggaacccctaat 346
Ps 181 ctccgagcagagtgctccgactcgtcgtgagtcgaagggccacagatggcccaacgat 240
Cc 347 ctccgaattaggtttccctgtgttaagtgttcataaaaggaataatcgagatggt 406
Cc 241 gctgaggaaacbtgtaactccgcctggaagacccagcgatgataccgaagatggtgta 300
Cc 407 gctcgtctcgtgcatgataaattgaacagcctgagatattccacagat----- 458
Sg 301 cagcgagctcgtggaatggacagagatggcgctgaaatccggtcccatctccat 360
Db 459 -atcaatctgtgaataagttaaataaattcaaggaggtgttcctcctgctcgat 517
Qy 361 caaatggcagcttgagtgctgctgagtcgagagccagcgagcgatcgacaatc 420
Db 518 agaccatggcagatcaattgacggaggcaattgacggagaaacccgatgacgaagc 577
Qy 421 ctctgagatcgagcagacgagagagacggaagtcggtgttcgcaaatatctcgga 480
Db 578 atctggtctgtgctcttatgtgtaaggggcaaatcaacatacggaagtcacatac 637
Qy 481 ctcaagcccgactggtctcacatggtgtggaacacgaagagcgatgttaccagtc 540
Db 638 aagaagattggtctcacaccgggtgggaagagaggaataatcttctcctatgtg 697
Qy 541 atcgaggacccaaacgaatttaactcctcgatgtaccaggtgtaattgaagtatt 600
Db 698 gacgaaggatctgacaagcatagattgttatctcctggtgaatcagagattatt 757
Qy 601 aattatgacctgtagaatgtgttaagaacagggcattcagttcggaacaaatcacga 660
Db 758 aattatgagtaataagcattaaagataagggttatagagagtaactaaatacaaac 817
Qy 661 cttagttattctgggttcgacatgctgagtgactcgtatttgccaatgctcgtg 720
Db 818 ataaagatagttgaattagtaaaatcacatgtaactgctgagcgaatttcagtcg 877
Qy 721 tatttgaataatcagcggagacagaataaaactgtggaatatttaagta 769
Db 878 ttctgtaaaatctccgaagatcgataaaatacaattatttctgctgaagaa 926

```

RESULT 7  
AAA38946  
ID AAA38946 standard; DNA; 1106 BP.

XX AC AAA38946;

XX DT 23-AUG-2000 (first entry)

XX DE Banana bunchy top virus related nucleotide sequence #1.

XX DE Banana bunchy top virus; BBTV; detection; ds.

XX DE Banana bunchy top virus.

XX DE TW360710-A.

XX DE 11-JUN-1999.

XX DE 30-JUN-1994; 94TW-0106105.

XX DE 30-JUN-1994; 94TW-0106105.

XX DE (BIOT-) DEV CENT BIOTECHNOLOGY.

XX DE Wu R, You L, Song T;

```

XX DR WPI; 2000-316145/27.
XX PT Two circular single-stranded DNAs associated with banana bunchy top
XX PT virus and detection of the virus
XX PS Claim 1; Page 1; 7pp; Chinese.
XX CC The present invention describes two circular single-stranded DNAs
XX CC associated with banana bunchy top virus (BBTV) and the encoded proteins.
XX CC Also described is a method of detecting the BBTV virus using the
XX CC polymerase chain reaction (PCR). The present sequence represents a
XX CC BBTV related nucleotide sequence from the present invention.
XX Sg Sequence 1106 BP; 335 A; 226 C; 256 G; 289 T; 0 other;

```

Query Match 17.4%; Score 136.2; DB 21; Length 1106;  
Best Local Similarity 50.6%; Pred. No. 6.7e-36;  
Matches 389; Conservative 0; Mismatches 368; Indels 12; Gaps 2;

```

Qy 1 gttgggacagaggttgacacttcaactggccaagacacactccaaggtatccacttg 60
Db 170 gtcggcagcaatcgctcgccacccggaagacacccctccaggtatctatccctg 229
Qy 61 aagaccggtcgcgactcgaaggtgaagactgttcttgggaatgacaggttaccctg 120
Db 230 aaaaagagaatccgctcggtggtgaagaagatggt---ttccggtgctcactgg 286
Qy 121 gagcgcacccgtggttcgcagacacagaatagagactactgttcgaaggaacgggtgctt 180
Db 287 gagattgcagaggaaacgacgagagaaattcgaagtactgttcgaaggaacccctaat 346
Qy 181 ctccgagcagagtcggcgactcgtcgtgagtcgaagggccacagatgtggcccaacgat 240
Db 347 ctcaaatagggtttctgtgttgaatggttcttaataaaaggaataatcggagatggt 406
Qy 241 gctgaggaacctgatgaactccgctggaagacccagcggtacaccgaagatcggttga 300
Db 407 gctcgttctcgtcgtcgtgaaattgacagcgtgagatatttcacagat----- 458
Qy 301 cagcgagctcgtggaatggacagagatggcgctgaaatccggtcccatctccat 360
Db 459 -atcaatctgtgaataagttaaataaattcaaggaggttcgttcctcgtcgat 517
Qy 361 cacaattggcagcttgaggtgctgctgacgagcagcgagcgagcgatcgacaatc 420
Db 518 agaccatggcagatcaattgacggaggcaattgacggagaaacccgatgacgaagc 577
Qy 421 ctctgagatcgagcagcagcagcgagagcggaagtcggtgttcgcaaatatctcgga 480
Db 578 atctggtctatggtccttatgtgtaaggggtaatacaacatacggaagtcactaatc 637
Qy 481 ctcaagcccgactggttctcacatggtgtggaacacagggcattcagttcggaacaa 540
Db 638 aagaagattggttctcacaccgggtgggaagagaggaataatcttctcctatgtg 697
Qy 541 atcgaggacccaaacgaatttaactcctcgatgtaccaggtgtaattgaagtatt 600
Db 698 gacgaaggatctgacaagcatagattgttatctcctggtgaatcagagattatt 757
Qy 601 aattatgacctgtagaatgtgttaagaacagggcattcagttcggaacaaatcacga 660
Db 758 aattatgagtaataagcattaaagataagggttatagagagtaactaaatacaaac 817
Qy 661 cttagttattctgggttcgacatgctgagtgactcgtatttgccaatgctcgtg 720
Db 818 ataaagatagttgaattagtaaaatcacatgtaactgctgagcgaatttcagtcg 877
Qy 721 tatttgaataatcagcggagacagaataaaactgtggaatatttaagta 769
Db 878 ttctgtaaaatctccgaagatcgataaaatacaattatttctgctgaagaa 926

```

| Query Match           | 13.8%           | Score 108;   | DB 17;     | Length 1022; |
|-----------------------|-----------------|--|------------|--------------|
| Best Local Similarity | 49.3%;          | Pred. No. 2.5e-26;   |            |              |
| Matches 378;          | Conservative 0; | Mismatches 370;  | Indels 18; | Gaps 3;      |
| QY                    | 1               | gttgggacagaggttgacacttcaactggcgaacgacacactccaaggattcatccacttg  | 60         |              |
|                       |                 |  |            |              |
| Db                    | 178             | gttgcggacagaaactgcaactactgacagaaacactccaggattgtatcgctc         | 237        |              |
| QY                    | 61              | aagaccggtcggcactgcgaaggtgaagactgttcttgggaatgacagattcaactcg     | 120        |              |
|                       |                 |  |            |              |
| Db                    | 238             | aagaacaaaattcgtcttgggtgattgaagagaataattggttaatc---gagctcaactgg | 294        |              |
| QY                    | 121             | gagccgaccctgggtctccgacgacacgaataagactactgtctgaaagacaggtgctt    | 180        |              |
|                       |                 |  |            |              |
| Db                    | 295             | gaatttcgagagcgacgcgaattccgactcgcgatttctgttaagaaacacctaat       | 354        |              |

|          |  |
|----------|--|
| RESULT   | 9  |
| AAT13165 |  |
| ID       | AAT13165 standard; DNA; 1017 BP.                         |
| XX       |  |
| AC       | AAT13165;  |
| XX       |  |
| DT       | 23-MAY-1996 (first entry)                                |
| XX       |  |
| SCSV     | segment 6.   |
| XX       |  |
| KW       | SCSV; promoter; transcription; transgenic plant; legume; |
| KW       | gene expression; crop improvement; ss.                   |
| XX       |  |
| OS       | Subterranean clover stunt virus isolate F.               |
| XX       |  |
| Key      | Location/Qualifiers                                      |
| FH       | 48...905   |
| CDS      | /tag= a  |
| FT       |  |
| FT       |  |
| XX       |  |
| PN       | WO9606932-Al.  |
| XX       |  |
| PD       | 07-MAR-1996.   |
| XX       |  |
| PF       | 30-AUG-1995; 95WO-AU00552.                               |
| XX       |  |
| PR       | 07-NOV-1994; 94AU-0009281.                               |
| PR       | 30-AUG-1994; 94AU-0007770.                               |
| XX       |  |
| PPA      | (CSIR ) COMMONWEALTH SCI & IND RES ORG.                  |
| PA       | (AUSU ) UNIV AUSTRALIAN NAT.                             |
| XX       |  |
| PI       | Boevink PC, Chu PWG, Keese PK, Khan RI, Larkin PJ;       |

|           |  |
|-----------|--|
| RESULT    | 10   |
| AAAX34686 |  |
| ID        | AAAX34686 standard; DNA; 593 BP.                                     |
| XX        |  |
| XX        | AAAX34686;   |
| AC        |  |
| XX        |  |
| XX        |  |
| DT        |  |
| XX        | 02-JUL-1999 (first entry)  |
| DE        | DNA sequence of BBTV SI promoter fragment including an ORF fragment. |
| XX        |  |
| KW        | promoter; Banana Bunchy Top Virus; BBTV; gene transcription; ubil;   |
| KW        | polyubiquitin 1; banana; ss.   |
| XX        |  |
| OS        | Banana bunchy top virus.   |
| XX        |  |
| PN        | W09915646-A1.  |

|     |   |     |
|-----|---|-----|
| 480 | actcaagcccgactggtttctacaactgtgttggaacacagaagaagctattgtaccacta     | 533 |
| QY  |   |     |
| 481 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 52  | aatgaagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga      | 111 |
| QY  |   |     |
| 53  | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 540 | cat---cgaggaccacaaacgaatttaactctcgatgtaccagatgtaatttagagta        | 596 |
| QY  |   |     |
| 541 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 112 | cgTgcagcaagatctgagaagcattgtatttgatatccctgctgtaatcaggatta          | 171 |
| QY  |   |     |
| 113 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 597 | tttaaatattgcctgttttagaatgtgtttaagaacacagaggaattcagttcggaacaaatcga | 656 |
| QY  |   |     |
| 598 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 172 | tttaaatattgattatagagcattaaagacagaggtgatagagagactacaataataa        | 231 |
| QY  |   |     |
| 173 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 657 | accccttagtattcttgggttcgcacatgtgcactgtactcgtatttgcgaatgctctgcc     | 716 |
| QY  |   |     |
| 658 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 232 | accttaataagttagttgaaattgatttaataacatgtaattgctggttaatttcattgcc     | 291 |
| QY  |   |     |
| 233 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 717 | tgattatttgaataatcagcagggacagaataaaactgtggaaattatttaa              | 766 |
| QY  |   |     |
| 718 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |
| 292 | agaattctgtLaaaatctccgaagatagataaagaattattattgtgtctaa              | 341 |
| QY  |   |     |
| 293 | aaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga        |     |
| DB  |   |     |

|          |   |
|----------|---|
| RESULT   | 11  |
| AAV24089 |   |
| ID       | AAV24089 standard; DNA; 1110 BP.                  |
| XX       |   |
| XX       |   |
| AC       | AAV24089;   |
| XX       |   |
| DT       | 11-AUG-1998 (first entry)                         |
| XX       |   |
| DE       | Banana bunchy top virus component 1 DNA sequence. |
| XX       |   |
| XX       |   |
| XX       | BTV; probe; diagnostic primer; component 1; ss.   |
| OS       |   |
| XX       | Banana bunchy top virus.                          |
| XX       |   |
| PN       | US5756708-A.                                      |
| XX       |   |
| XX       |   |

|   |  |               |
|---|--|---------------|
| PF  | 31-MAY-1996;   | 96WO-AU00335. |
| XX  |  |               |
| PR  | 31-MAY-1995;   | 95AU-0003285. |
| XX  |  |               |
| PA  | (UYQU-) UNIV QUEENSLAND TECHNOLOGY.                                  |               |
| XX  |  |               |
| PI  | Becker DK, Beetham PR, Dale JL, Dugdale B, Hafner GJ;                |               |
| XX  | Harding RM;  |               |
| DR  | WPI; 1997-034368/03.   |               |
| XX  |  |               |
| PT  | DNA from intergenic region of banana bunchy top virus DNA component  |               |
| PT  | - useful for promoting, enhancing, regulating or modifying           |               |
| PT  | transcription of a non-BBTV gene in transgenic plants                |               |
| XX  |  |               |
| PS  | Claim 4; Fig 12; 80pp; English.                                      |               |
| XX  |  |               |
| CC  | Intergenic regions (AAT49399-409) are derived from genomic DNA       |               |
| CC  | components 1-6 (see also AAT49386-90) of banana bunchy top virus     |               |
| CC  | (BBTV). A large intergenic region (AAT49405) comprises the insert    |               |
| CC  | in plasmid pBR1.INT, a plasmid obtd. by PCR amplification            |               |
| CC  | (see also AAT49427-28) of a full-length BBTV clone and cloning of    |               |
| CC  | tagI-digested PCR product into pBl101.3. BBTV intergenic regions     |               |
| CC  | are useful for promoting, enhancing, regulating or modifying         |               |
| CC  | transcription of non-BBTV genes in monocotyledon or dicotyledon      |               |
| CC  | transgenic plants. The non-BBTV gene is e.g. an insecticide          |               |
| CC  | resistance gene, herbicide resistance gene or growth promoting       |               |
| CC  | gene.  |               |
| XX  |  |               |
| SQ  | Sequence 982 BP; 316 A; 174 C; 249 G; 243 T; 0 other;                |               |
| <br>  |  |               |
| Query Match            7.6%; Score 59.6; DB 18; Length 982; |  |               |
| Best Local Similarity 51.3%; Pred. No. 7e-10;               |  |               |
| Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps |  |               |
| 2;  |  |               |
| QY  | 410 atcgcaaatcctctggatgacgcagcagcagcagggaagtccgtgttgcca 469          |               |
| Db  |  |               |
| QY  | 615 atcgagaattatttgggttacgccaaatggagagagaaagacaacgatgcaa 674         |               |
| QY  | 470 aatatctcgactcaagccgcagactgtttcacatgtggtgaaaccagaaaggacgtat 529   |               |
| Db  |  |               |
| QY  | 675 aacataatgaagcagaagaatgcgtttttattcttcaggaggaaaatcatgtgatata 734   |               |
| QY  | 530 tgtaccagtaacatcgaggccccaaacgaatttaacctcgtatgcccaaggttaatt 589    |               |
| Db  |  |               |
| QY  | 735 gttagctgtataattagggatat-----tgttatattgatcccctagatgcaag 788       |               |
| QY  | 590 tagagtatttaaattatgcctgttagaattgtttaagacacagggcattcagttcggaca 649 |               |
| Db  |  |               |
| QY  | 789 aggattatttaaattatgtttattagaggaattlaagaatggaataattcaaacgggga 848  |               |
| QY  | 650 aatacaaaccccttaattatttgggttcgacatgtgcgatctactgtatttgccaatg 709   |               |
| Db  |  |               |
| QY  | 849 aatatgaacccgtttgaagattg---tagaataatgcgaadtcattgtaattgctaact 905  |               |
| QY  | 710 tctgcctgattatttgaataatcagcagggacagataaaactg 753                  |               |
| Db  |  |               |
| Db  | tccctccgaaggagaactctttctgaagatcgaaataaagtgtg 949                     |               |
| <br>  |  |               |
| RESULT 12   |  |               |
| ID  | AAT49405   |               |
| DE  | AAT49405 standard; DNA; 982 BP.                                      |               |
| AC  | AAT49405;  |               |
| XX  |  |               |
| DT  | 23-AUG-1997 (first entry)  |               |
| XX  |  |               |
| DE  | Banana bunchy top virus DNA intergenic region 1 insert in pBR1.INT.  |               |
| KW  | BBTV; intergenic region; promoter; transgenic plant; ds.             |               |
| OS  | Banana bunchy top virus.   |               |
| XX  |  |               |
| PN  | WO9638554-A1.  |               |
| PD  | 05-DEC-1996.   |               |
| XX  |  |               |

DE Banana bunchy top virus component 1 DNA sequence.  
XX  
XX  
KW BBTv; probe; diagnostic primer; component 1; ss.  
XX  
XX  
OS Banana bunchy top virus.

Query Match 7.6%; Score 59.6; DB 19; Length 1110;  
Best Local Similarity 51.3%; Pred. No. 7.6e-10;  
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

[illegible]

RESULT 14  
AAV24077  
ID AAV24077 standard; DNA; 1111 BP.  
XX  
XX AAV24077;  
XX

|    |  |                                     |
|----|--|-------------------------------------|
| DT | 11-AUG-1998  | (first entry)                       |
| XX | Banana bunchy top virus component 1                                      | DNA sequence.                       |
| DE | Banana bunchy top virus component 1                                      | DNA sequence.                       |
| XX | BBTV; probe;   | diagnostic primer; component 1; ss. |
| XX | Banana bunchy top virus.   |                                     |
| OS | Key  | Location/Qualifiers                 |
| XX | 129..989   |                                     |
| FT | CDS  | /*tag= a                            |
| FT |  |                                     |
| XX | US5756708-A.   |                                     |
| PN |  |                                     |
| XX | 26-MAY-1998.   |                                     |
| PD |  |                                     |
| XX | 24-FEB-1994;   | 94US-0202186.                       |
| PF |  |                                     |
| XX | 24-FEB-1994;   | 94US-0202186.                       |
| XX |  |                                     |
| PR | (UYQU-) UNIV QUEENSLAND TECHNOLOGY.                                      |                                     |
| XX |  |                                     |
| XX | Burns TM, Dale JL, Harding RM, Karan M;                                  |                                     |
| PA |  |                                     |
| PI |  |                                     |
| XX | WPI; 1998-321636/28.   |                                     |
| XX | P-PSDB; AAW54071.  |                                     |
| DR |  |                                     |
| XX | Isolated Banana bunchy top virus DNA                                     | - useful as diagnostic probes       |
| PT | and primers and for producing virus-resistant plants                     |                                     |
| XX | Example 1; Fig 11; 59pp; English.  |                                     |
| XX | This sequence represents a DNA fragment from the component 1 of          |                                     |
| CC | the Banana bunchy top virus (BBTV). The DNA molecules can be used as     |                                     |
| CC | diagnostic probes or primers or can be inserted into plants or other     |                                     |
| CC | organisms e.g. to produce virus-resistant plants or to act as promoters, |                                     |
| CC | enhancers or termination signals.  |                                     |
| XX | Sequence 1111 BP; 363 A; 192 C; 276 G; 280 T; 0 other;                   |                                     |
| XX |  |                                     |
| SO |  |                                     |

|                       |              |                    |                 |              |
|-----------------------|--------------|--------------------|-----------------|--------------|
| Query Match           | 7.6%;        | Score 59.6;        | DB 19;          | Length 1111; |
| Best Local Similarity | 51.3%;       | Pred. No. 7.6e-10; |                 |              |
| Matches 193;          | Conservative | 0;                 | Mismatches 174; | Indels 9;    |
| Gaps                  |              |                    |                 |              |

|    |     |   |     |
|----|-----|---|-----|
| Qy | 378 | agtgctgtctgcgacgcgagagccagcgagcgcgcacacatctctcggtatcgcgacg      | 437 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 611 | agaggtggagaatacatcgcgcgccatgtcatcgagagataaattgggtctatggccc      | 670 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Qy | 438 | agacggagagacgggaagtcctgtttgcacaaatactcggactcaagcccgactggtt      | 497 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 671 | aaatggagagaggaagacacacgtatgcaaacatctaatgaagcagagaaatgcgtt       | 730 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Qy | 498 | ctacacatgtggtggaaaccgagaagagcgctattgtaccagtcacatcgaggaccacaaacg | 557 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 731 | ttattctccaggagagaaatcattggatatactagactgtataattacgaggatat---     | 787 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Qy | 558 | aaatttaactcctgattgaccaggtgtaatttagagtatttaattatgcctgttgtaga     | 617 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 788 | ---tggtatttgatatctcaagatgcaaaagagattatttaattatgggttatttaga      | 844 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Qy | 618 | atgtgtttaagaacagggaattcagttcggacaatacgaaccccttagttatcttgggtt    | 677 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 845 | ggaatttaagaataaggaataatccaagcgggaataatgaacccgttttgaagatag---    | 901 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Qy | 678 | cgaccatgtgatgtactcgtatttgcgaatgtcctgcctgattatttgaaaatcagcag     | 737 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 902 | agaatatgtcgaagcaattgttaatggctaaactcctccgaaggaagaatctttctcga     | 961 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Qy | 738 | ggacagaataaaaactg   | 753 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 962 | agatcgaataaagtgtg   | 977 |



Search completed: October 24, 2001, 10:07:20  
Job time: 625 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 09:56:55 ; Search time 393.52 Seconds  
(without alignments)  
2059.923 Million cell updates/sec

Title: US-09-462-955-1

Perfect score: 1291

Sequence: 1 cgcacaaacacctctgtaag.....ggacggctgagtgtgctgg 1291

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 179.2 | 13.9        | 1096   | 20 | AAV71833    |
| 2          | 179.2 | 13.9        | 1096   | 21 | AAV38950    |
| 3          | 163.4 | 12.7        | 1106   | 20 | AAV71832    |
| 4          | 161.8 | 12.5        | 1106   | 21 | AAV38946    |
| 5          | 159.6 | 12.4        | 1091   | 20 | AAV71834    |
| 6          | 159.6 | 12.4        | 1091   | 21 | AAV38947    |
| 7          | 145.4 | 11.3        | 886    | 20 | AAV34687    |
| 8          | 112.4 | 8.7         | 1022   | 17 | AAT13161    |
| 9          | 99.2  | 7.7         | 1017   | 17 | AAT13165    |
| 10         | 81.2  | 6.3         | 593    | 20 | AAV34686    |
| 11         | 62.8  | 4.9         | 1110   | 19 | AAV24089    |

|    |      |     |      |    |           |                    |
|----|------|-----|------|----|-----------|--------------------|
| 12 | 59.6 | 4.6 | 982  | 18 | AAV49405  | Banana bunchy top  |
| 13 | 59.6 | 4.6 | 1110 | 19 | AAV24086  | Banana bunchy top  |
| 14 | 59.6 | 4.6 | 1111 | 19 | AAV24077  | Banana bunchy top  |
| 15 | 59.6 | 4.6 | 1111 | 19 | AAV24084  | Banana bunchy top  |
| 16 | 59.6 | 4.6 | 1111 | 19 | AAV24087  | Banana bunchy top  |
| 17 | 58.6 | 4.5 | 1103 | 19 | AAV24091  | Banana bunchy top  |
| 18 | 58.6 | 4.5 | 1105 | 19 | AAV24093  | Banana bunchy top  |
| 19 | 58   | 4.5 | 1109 | 19 | AAV24088  | Banana bunchy top  |
| 20 | 58   | 4.5 | 1111 | 19 | AAV24090  | Banana bunchy top  |
| 21 | 57   | 4.4 | 1104 | 19 | AAV24092  | Banana bunchy top  |
| 22 | 56.4 | 4.4 | 1111 | 19 | AAV24085  | Banana bunchy top  |
| 23 | 49.4 | 3.8 | 936  | 22 | AAF58252  | Oligonucleotide D1 |
| 24 | 49.4 | 3.8 | 936  | 22 | AAF58252  | Oligonucleotide D1 |
| 25 | 49.4 | 3.8 | 936  | 22 | AAF58254  | Oligonucleotide D1 |
| 26 | 49.4 | 3.8 | 936  | 22 | AAF58254  | Oligonucleotide D1 |
| 27 | 49.4 | 3.8 | 936  | 22 | AAF58257  | Oligonucleotide D1 |
| 28 | 49.4 | 3.8 | 936  | 22 | AAF58257  | Oligonucleotide D1 |
| 29 | 49.4 | 3.8 | 936  | 22 | AAF58259  | Oligonucleotide D1 |
| 30 | 49.4 | 3.8 | 936  | 22 | AAF58259  | Oligonucleotide D2 |
| 31 | 49.4 | 3.8 | 936  | 22 | AAF58262  | Oligonucleotide D2 |
| 32 | 49.4 | 3.8 | 936  | 22 | AAF58262  | Oligonucleotide D2 |
| 33 | 49.4 | 3.8 | 938  | 22 | AAF58255  | Oligonucleotide D1 |
| 34 | 49.4 | 3.8 | 938  | 22 | AAF58255  | Oligonucleotide D1 |
| 35 | 46.4 | 3.6 | 287  | 20 | AAV71830  | Subgenomic fragmen |
| 36 | 46.4 | 3.6 | 287  | 21 | AAV38948  | Banana bunchy top  |
| 37 | 46.4 | 3.6 | 300  | 20 | AAV71831  | Subgenomic fragmen |
| 38 | 46.4 | 3.6 | 300  | 21 | AAV38949  | Banana bunchy top  |
| 39 | 38   | 2.9 | 244  | 22 | AAF58238  | Oligonucleotide D1 |
| 40 | 38   | 2.9 | 244  | 22 | AAF58238  | Oligonucleotide D1 |
| 41 | 36.4 | 2.8 | 1939 | 21 | AAV29647  | Human choline/etha |
| 42 | 33   | 2.6 | 9345 | 21 | AAV293966 | Mouse uromodulin p |
| 43 | 32.8 | 2.5 | 1768 | 20 | AAV83754  | Porcine circovirus |
| 44 | 32.6 | 2.5 | 5053 | 19 | AAV22816  | Nucleotide sequenc |
| 45 | 32.6 | 2.5 | 5055 | 19 | AAV17236  | DNA from a region  |

## ALIGNMENTS

RESULT 1  
ID AAV71833 standard; DNA; 1096 BP.  
XX AC AAV71833;  
XX AC AAV71833;  
DT 10-FEB-1999 (first entry)  
XX BBTV DNA II clone (2-17) nucleotide sequence.  
DE BBTV DNA II clone (2-17) nucleotide sequence.  
XX BBTV DNA II clone (2-17) nucleotide sequence.  
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
XX Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
OS Banana bunchy top virus.  
XX Banana bunchy top virus.  
FH Key Location/Qualifiers  
FT TATA\_signal 1..7  
FT stem\_loop 8..38  
FT CDS 70..928  
FT polyA\_signal 1030..1035  
FT polyA\_signal 1030..1035  
XX US5846705-A.



RESULT 3  
AAV71832  
ID AAV718  
XX



|          |   |   |     |
|----------|---|---|-----|
| Dd       |   | 690 cctatgtggcgaaggatctgcacagcatatagattttgatcttccttctaactccagg        | 749 |
| OY       |   | 803 agcatttaaatatgccctgttagaaatgtgttaaacaaggcgattcagttcggacaatat      | 862 |
| Dd       |   | 750 attatttaaatatgatagaaggcataaaggatagggttatagagagatactaataat         | 809 |
| OY       |   | 863 acgaaccaccttgattctatcttggttgtcacocatgtgcattactctgtatttgccaatgtcc  | 922 |
| Dd       |   | 810 accaacccataaagatagttgaattaggttaaaaatacatgtaaatogtcaatgcccgaattcca | 869 |
| OY       |   | 923 tgcctgattattgaaaaatcagcagggacagacaataaaaactgtggaatattttaagata     | 979 |
| Dd       |   | 870 tgcctgactctgtaaaatctccccgaagatcgataaaaaatcattttatigtgaagaa        | 926 |
| RESULT   | 5   |   |     |
| AAV71834 |   |   |     |
| ID       | AAV71834 standard; DNA; 1091 BP.  |   |     |
| XX       |   |   |     |
| AC       | AAV71834;   |   |     |
| XX       |   |   |     |
| DT       | 10-FEB-1999 (first entry)   |   |     |
| XX       |   |   |     |
| DE       | BTV DNA II clone (2) nucleotide sequence.                                 |   |     |
| KW       | Banana bunchy top virus; BBTv; PCR technique; plant; tissue; banana;      |   |     |
| KW       | Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss. |   |     |
| OS       | Banana bunchy top virus.  |   |     |
| XX       |   |   |     |
| FN       | US5846705-A.  |   |     |
| PD       | 08-DEC-1998.  |   |     |
| XX       |   |   |     |
| PP       | 06-APR-1995; 95US-0418071.  |   |     |
| XX       |   |   |     |
| PR       | 06-APR-1995; 95US-0418071.  |   |     |
| XX       | (BIOT-) DEV CENT BIOTECHNOLOGY.   |   |     |
| PA       |   |   |     |
| XX       | Soong T, Wu R, You L;   |   |     |
| PI       | WPL; 1999-059037/05.  |   |     |
| DR       | Nucleic acids having banana bunchy top virus component sequences -        |   |     |
| XX       | used to design primers for use in polymerase chain reaction               |   |     |
| PT       | detection of the virus  |   |     |
| XX       |   |   |     |
| PS       | Disclosure; Fig 12A-C; 27pp; English.                                     |   |     |
| CC       | This represents the nucleotide sequence of a banana bunchy top virus      |   |     |
| CC       | ((BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention   |   |     |
| CC       | provides nucleic acid sequences associated with BBTV that can be used in  |   |     |
| CC       | a PCR technique for detecting BBTV. The nucleic acid sequences ((AAV71830 |   |     |
| CC       | to AAV71833)) are used as the basis for the construction of PCR primers,  |   |     |
| CC       | to detect BBTV infection. The PCR technique is used for detecting BBTV    |   |     |
| CC       | in plant tissues (preferably banana, especially Musa species). The        |   |     |
| CC       | virus, one of the most important banana species viruses, causes phloeom   |   |     |
| CC       | damage and is transmitted by aphids. PCR detection gives accurate,        |   |     |
| CC       | reliable and specific determination of absence or presence of the         |   |     |
| CC       | virus.  |   |     |
| SQ       | Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other:                    |   |     |

|          |   |  |     |
|----------|---|--|-----|
| QY       | 126   | gtgcttcaactctgaactcagacagagaaagagcgcgcaaatgttgttaagcgcgcatcga  | 185 |
| Db       | 88  | gtgcttcaactctgaatattctctcgacgagcgagaaagactttctcgctctctcgaa     | 147 |
| QY       | 186   | gtcccttaacttgctctatgtctatagtttggggagcaggttgcaccttcaactggccaacg | 245 |
| Db       | 148   | ggaggaaagagttcaattacgctgtcgttcggcagcagaagtgcgtccgagcaccggtcgaa | 207 |
| QY       | 246   | acacctcaagagattccacttgcattgaagacgcgtcgcgactgcaagagattgaagactgt | 305 |
| Db       | 208   | gcaacctcagggatctatccctgaagaaatctttaagctgtgtgattgaagagagag      | 267 |
| QY       | 306   | tcttgggaatgacaagattcaacttggagccaccgcgtgttccgcagcaacagatatagaa  | 365 |
| Db       | 268   | gtactc---tctgaaggtcactggagagggcgagaggaactgatgaacagaatccgag     | 324 |
| QY       | 366   | ctactgttcgaaggaaacgggtgcttcttcgacgacgagtcgccgactcgtctctgagtcac | 425 |
| Db       | 325   | atactgttcgaaggaa-----acccttgaaactgggtactcccggttctcgtgttcgaa    | 379 |
| QY       | 426   | aaggccacgattggccaaacgatttgtcggagaaactcgtatgaaactccgctcgagaaacc | 485 |
| Db       | 380   | gaagcgcaagcttctcgatgattcagagagagccctgaggaattgaagatggacgatcc    | 439 |
| QY       | 486   | agcgcgataccgaagatgcgtgttcacgcgagcttcggttggaatgacaagaatggccgc   | 545 |
| Db       | 440   | atccaagtatcgagatgcttggcagtggaatcaattaaagatgcagaaatlaattccga    | 499 |
| QY       | 546   | tgaataatcgttccattccatcatcacattggcagcttgaagctgtctctcgcgacg      | 605 |
| Db       | 500   | atgggttcgaaacta-----aaagaatggcaaaataaattaatccaacatcga          | 550 |
| QY       | 606   | agagccagcgacgatcgacaatcctctgatatcggaacgagacgagggagacgggaa      | 665 |
| Db       | 551   | aggtgtcctgatgatcgaagtaaccactdgggtatcgcgtcctgcggagggcgaaggaa    | 610 |
| QY       | 666   | gtccggttgccaaatatctcgaactcaagcccgactggttctcacactgttgggtggaac   | 725 |
| Db       | 611   | gtcaaccttcgaagatctatcattaaaaacctggatgggatatacaacggtggaaa       | 670 |
| QY       | 726   | cagaaagagctattgtaccagttacatcgaggaccaccaaaacgaaatttaactcctgatgt | 785 |
| Db       | 671   | gactcgggatgatgacacatcaacgatggatggatcctgatactgattgatttgatat     | 730 |
| QY       | 786   | accaggtgtaatttagtatatttaaatattgcctgttgagaatgtgttaagaacagagc    | 845 |
| Db       | 731   | ccccagaagtcattcagattatctgaattatggcgttatagaacaaattcaagaatagagt  | 790 |
| QY       | 846   | attcagttcggacaaatacgaaccc---ccttagttattcttgggttcgacctatgcatgt  | 902 |
| Db       | 791   | tttaataaaatacaaaatacgaacacatgtgtgattagaagaatggacaaaatgtccatgt  | 850 |
| QY       | 903   | actcgtatttgcgaatgctcgtcctgatatttggaaatccagcagggagacagaataaaact | 962 |
| Db       | 851   | aattgttatggcaaatgtgttcgctgatttattgtaaaatttcagaagatagaataaaaat  | 910 |
| QY       | 963   | gtggaatatttaagtatgtgtcacc                                      | 988 |
| Db       | 911   | aattaatgttgaaaggaaacttc  | 936 |
| RESULT 6 |   |  |     |
| AAA38947 |   |  |     |
| ID       | AAA38947 standard; DNA; 1091 BP.                        |  |     |
| XX       | AAA38947;   |  |     |
| AC       | AC  |  |     |
| XX       | AC  |  |     |
| DT       | 23-AUG-2000 (first entry)                               |  |     |
| XX       | Banana bunchy top virus related nucleotide sequence #2. |  |     |
| DE       | Banana bunchy top virus; BBTV; detection; ds.           |  |     |
| XX       |   |  |     |
| KW       |   |  |     |
| XX       |   |  |     |





[illegible]











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 09:56:55 ; Search time 5701.85 Seconds  
(without alignments)  
3502.173 Million cell updates/sec

Title: US-09-462-955-1  
Perfect score: 1291  
Sequence: 1 ccaccaaaacctgctgaag.....ggacggcgtgagtgcctgg 1291

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_ba3:  
4: gb\_in1:  
5: gb\_in2:  
6: gb\_in3:  
7: gb\_om:  
8: gb\_ov:  
9: gb\_pat1:  
10: gb\_pat2:  
11: gb\_ph:  
12: gb\_pl1:  
13: gb\_pl2:  
14: gb\_pl3:  
15: gb\_pl4:  
16: em\_ba1:  
17: em\_ba2:  
18: em\_fun:  
19: em\_htgo\_hum:  
20: em\_htgo\_inv:  
21: em\_htgo\_rod:  
22: em\_htg\_hum1:  
23: em\_htg\_hum2:  
24: em\_htg\_hum3:  
25: em\_htg\_hum4:  
26: em\_htg\_hum5:  
27: em\_htg\_hum6:  
28: em\_htg\_hum7:  
29: em\_htg\_hum8:  
30: em\_htg\_inv1:  
31: em\_htg\_inv2:  
32: em\_htg\_other:  
33: em\_htg\_rod:  
34: em\_hum1:  
35: em\_hum2:  
36: em\_hum3:  
37: em\_hum4:  
38: em\_hum5:  
39: em\_hum6:  
40: em\_hum7:  
41: em\_in:  
42: em\_om:  
43: em\_or:

44: em\_ov:  
45: em\_pat:  
46: em\_ph:  
47: em\_pl:  
48: em\_ro:  
49: em\_sts:  
50: em\_sy:  
51: em\_un:  
52: em\_vi:  
53: gb\_sts1:  
54: gb\_sts2:  
55: gb\_sts3:  
56: gb\_sy:  
57: gb\_un:  
58: gb\_vil:  
59: gb\_vil2:  
60: gb\_vil3:  
61: gb\_vil4:  
62: gb\_vil5:  
63: gb\_vil6:  
64: gb\_vil7:  
65: gb\_vil8:  
66: gb\_vil9:  
67: gb\_vil10:  
68: gb\_vil11:  
69: gb\_vil12:  
70: gb\_vil13:  
71: gb\_vil14:  
72: gb\_vil15:  
73: gb\_vil16:  
74: gb\_vil17:  
75: gb\_vil18:  
76: gb\_vil19:  
77: gb\_vil20:  
78: gb\_vil21:  
79: gb\_vil22:  
80: gb\_vil23:  
81: gb\_vil24:  
82: gb\_vil25:  
83: gb\_vil26:  
84: gb\_vil27:  
85: gb\_vil28:  
86: gb\_vil29:  
87: gb\_vil30:  
88: gb\_vil31:  
89: gb\_vil32:  
90: gb\_vil33:  
91: gb\_vil34:  
92: gb\_vil35:  
93: gb\_vil36:  
94: gb\_vil37:  
95: gb\_vil38:  
96: gb\_vil39:  
97: gb\_vil40:  
98: em\_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 1291  | 100.0       | 1291   | 58    | CFDCG       |
| 2          | 179.2 | 13.9        | 1096   | 9     | AR063452    |
| 3          | 179.2 | 13.9        | 1096   | 58    | BYTV2       |
| 4          | 173   | 13.4        | 1095   | 58    | AF216222    |
| 5          | 163.4 | 12.7        | 1106   | 9     | AR063451    |
| 6          | 163.4 | 12.7        | 1106   | 58    | BYTV1       |
| 7          | 160.8 | 12.5        | 1111   | 58    | BBU12586    |
| 8          | 160.4 | 12.4        | 1127   | 58    | BBU12587    |

M29963 Coconut fol  
AR063452 Sequence  
L32167 Banana bunc  
AF216222 Banana bu  
AR063451 Sequence  
L32166 Banana bunc  
U12586 Banana bunc  
U12587 Banana bunc

AR063453 Sequence  
AU1316221 Banana bu  
U02312 Banana bunc  
AU000921 Milk vetc  
AU7005961 faba bean  
AU7005962 faba bean  
AU132187 faba bean  
AU7005968 faba bean  
AU16731 Subterranea  
UB039047 Milk vetc  
AU132185 faba bean  
AU7005964 faba bean  
AU8000922 Milk vetc  
U16735 Subterranea  
X80879 FBNYV gene  
AU000920 Milk vetc  
AU132345 Nanovirus  
AU238493 Nanovirus  
AU132344 Nanovirus  
AU10230 Sequence  
AR112007 Sequence  
AU100227 Sequence  
U18077 Banana bunc  
AU102325 Sequence  
AU10228 Sequence  
AR10240 Sequence  
AU102780 Banana bu  
S56276 orf VI com  
AU10232 Sequence  
AR102324 Sequence  
AU10229 Sequence  
AU10231 Sequence  
AR123874 Banana bu  
U97525 Banana bunc  
AU102233 Sequence  
AR102226 Sequence  
AF238875 Banana bu  
AF246123 Banana bu

```

/ud_xref=GI:253307
/translation="MGSSIRRWCFLLNYETEEEAANVVRRRIESLNLVYAIVGDEVAPS
TGGRHLOGFIHIKTRGRRLGKTVLGNDRIHLEPTRGSDQNRDYSCKERVLLLEHGVP

```



```

Db 361 AGAGACTACTGTCGAGCAACGGGTCTCTCCAGACGAGTCCCGACTCGCTGGA 420
Qy 421 gtcaaaagccacgattgcccacgattgctgaggaacctgatgaactccgcctggaa 480
Db 421 GTCAAAAGCCACGATGCGCCCAACGATTTGCTGAGGAACCTGATAACTCCGCCCTGGAA 480
Qy 481 gaccagcggtatccacgaagatgctgtacacgagcttcgctgaatggacaagatgg 540
Db 481 GACCCAGCGCATACCGAAGATGGTGTACACGAGCTTCGGTGGATGGCAAGATGG 540
Qy 541 gccctgaaatccgttccattccatatacaaatggagcttgaagtgtctgtcg 600
Db 541 GCCCTGAAATCCGTTCCTCATATCCATATCAAAATGGCAGCTTCAAGTGTGCTGG 600
Qy 601 atcgagagccagcgagcagtcacaaactctctgataatcggaacgagacgagagac 660
Db 601 ATCGAGAGCCAGCGGACGATGCGCAATATCTCGGACTCAAGCCCGACTGGTCTACACATGTGGT 660
Qy 661 ggaagtcggtgttgccaaatatctcgactcaagcccgactggtttcacatgtggt 720
Db 661 GGAAGTCCGTGTTGCAAAATATCTCGGACTCAAGCCCGACTGGTCTACACATGTGGT 720
Qy 721 ggaaccagaagacgattgtaccagatcagaccccttagtattctgggttcgacctc 780
Db 721 GGAACCAAGAGACGATGTTACCAAGTACATCGAGAGCCCAAAACGAAATTTAATCTCT 780
Qy 781 gatgacccaggtgtaattagtagtattaaattatgcctcttagaagtgttaagac 840
Db 781 GATGTATCCCAAGTGTAAATTAGAGTATTTAAATTTATGCCCTGTAGATGTTAAGAAC 840
Qy 841 agggcattcagtcgacaaatcagaccccttagtattctgggttcgacctcgtgcat 900
Db 841 AGGGCATTCAGTTCGACAAATACGAACCCCTTAGTATTCTTGGGTTTCGACATGTGCAT 900
Qy 901 gtactgatttgccaatgctgctgattatttgaaatcagcagagcagataaaaa 960
Db 901 GTACTCGTATTTGCCAATGCTGCTGATATTTGAAATACAGAGGAGAGAATAAAA 960
Qy 961 ctgtgaatatataagtagtgctcatctaaattacacccatgacccgcccacacgc 1020
Db 961 CTGTGGAATATTTAAAGTATGTGTCTATCTAAATACACCAATACCCGCCGCCACCGC 1020
Qy 1021 tatcgtttacatttatgaatactctgcccagcgagcagcagcagcagcagcagc 1080
Db 1021 TATCGTTTACATCTTATGAAATATCGAGTTATGCGCGGCGGAGGCTGGAGGCTACCGCGC 1080
Qy 1081 cgaagcgggggaacaatatgaaatcgagttatggcgggcccaataaaagattccatt 1140
Db 1081 CGAAGCGCGGGAACAATATGAATCGAGTTATGCGCGGCGGCGGAGGCTGGAGGCTACCGCGC 1140
Qy 1141 ggataagaacaaatctgttacttctgctgagtcgagtcgacgcaacacacttccacgaccaa 1200
Db 1141 GGATGAAGAACAAATCTGTACTTTGCTGTGAGTGCAGCGCAACCACTTCCACGCTACCAA 1200
Qy 1201 tccaggtgagtagctgtgagagaagcgcgcaagcctctctacgctcattgtt 1260
Db 1201 TCCAGGTGAGTAGCTGTGAGAGAAGAAAGCCGCAAGCCCTCTATACCGTCCATTGTT 1260
Qy 1261 tttagcagatcgagcgtgagttgactgtg 1291
Db 1261 TTTGGCGATCGGAGCGGCTGAGTTGATCTGG 1291

```

```

RESULT 2
AR063452 1096 bp DNA PAT 29-SEP-1999
LOCUS
DEFINITION Sequence 4 from patent US 5846705.
ACCESSION AR063452
VERSION AR063452.1 GI:5992760
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

```

```

Unclassified.
1 (bases 1 to 1096)
Wu, R., You, L. and Soong, T.
Nucleotide sequence of two circular SSDNA associated with banana
bunchy top virus and method for detection of banana bunchy top
virus
Patent: US 5846705-A 4 08-DEC-1998;
Location/Qualifiers
1..1096
/organism="unknown"
BASE COUNT 347 a 231 c 244 g 274 t
ORIGIN

Query Match 13.9%; Score 179.2; DB 9; Length 1096;
Best Local Similarity 52.4%; Pred. No. 1.6e-42;
Matches 499; Conservative 0; Mismatches 433; Indels 20; Gaps 4;

Qy 40 agccgcggttagtattaccgccggtcccaacctgtgtaacccgcttggtta 99
Db 7 AGGAGGAGCGGCTAGTATTACCGGCTCCCTCGCACATTCCTCTCGCACCTGACGTC-- 64
Qy 100 taaatgggttccctccattccgcgctgcttcaactctgaactcagacagagaaag 159
Db 65 ---ATCATATTGCTCTCTTTAAATGGTGTCTCACTCTGAATATTCTCTCGCGCAGCGGAG 121
Qy 160 ggcgcgaattgttgaaggcgcatcgactcccttaacttgggtctctgtctatagttgggac 219
Db 122 CGAAGAAGACTTTCTCGCTCTTCTGAAGGAGGAGGATGTTCACTACTCTCTCTGCTGGCGAC 181
Qy 220 gaggttgcaccttcaactgccaacgacacactcccaagattccacttgaagaccggt 279
Db 182 GAAGTCCCTCCGGCCACCGCCAGAGCACTCCAGGGATATCTATCCCTGAAAAATCA 241
Qy 280 cggcgactgcaaggattgaagactgttcttggaatgacagattccactgagaccgacc 339
Db 242 ATTGCGCTCGCGGATTGAAAAGAAGTATGG---CTCCCGTGTCTACTGGGAGATTGCG 298
Qy 340 cgtgttcgcagcaagaatagagactactgttcgaaggacgggtgcttctcgcagcac 399
Db 299 AAAGGAAGTGACGACAACTCGCAGACTGTTCGAAGGAAACCCCTAGTCTTGAACCTG 358
Qy 400 ggagtcgccgactcgtcctggagtcacaaaggccacgattggcccaacgatttctgagaa 459
Db 359 GGTACTCCGGTGTCTCTGTTGTTGAAGAGCGCAAGTCTTCGATAGATTACAGAGAGC 418
Qy 460 cctgtagaactccgctggaagaccgagcggaataccgaagatgctgtgtacacgagct 519
Db 419 CCGTGAAGATTGAAGATGGACGATCCATCCAAAGTATCGCAGATGCTTGGCAGTGGAAATCA 478
Qy 520 ccggtggaatgacaaagatgggcgctgaaatccgttcccatttccatatacaaatgg 579
Db 479 ATTAAGATGCCAGAAATTAATTCGAATGGGTTCACCAACTA-----AAGAATGG 529
Qy 580 cagcttgaagtgtgtctgctgagagccgagcagcagcagcagcagcagcagcagcagc 639
Db 530 CAAATAAAATTAATTCACACATCGAAGGTGTTCTTGATGATGATGATGATGATGATGATGAT 589
Qy 640 tgcgagcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 699
Db 590 TACGGTCCCAACGAGCGGAGGAAAGTCAACCTTCGAAGATATCTATCAITATAAACCC 649
Qy 700 gactggttctacacatgtgtggaaccagaaagacgagcagcagcagcagcagcagcagc 759
Db 650 GGATGGGATATATCAACCGGTGGAAAGACGTCGGATATGATGATGATGATGATGATGATGAT 709
Qy 760 ccaaaacgaatttaactctcgtatgaccaggtgtaatttagagattatttaattatgcc 819
Db 710 CCGTAAATCAATTCGATATTGATATCCCGAGAGTCAATTCAGATTATCTGAATTATGATG 769
Qy 820 ctgttagaagtggttaagaacagcagcagcagcagcagcagcagcagcagcagcagcagc 876
Db 770 GTTATAGACAAATTAAGATAGAGTTTAAATAAATAACAAATACGAACCATGCTGATT 829

```











| Qy                    | 963          | gtggaatatttaagaagtatgtgcatc  | 988                 |
|-----------------------|--------------|--|---------------------|
|                       |              |  |                     |
| Db                    | 911          | ANTTAATTGTTGAGAAAGGAACCTC  | 936                 |
| RESULT                | 10           |  |                     |
| LOCUS                 | AF216221     |  |                     |
| DEFINITION            | AF216221     | 1109 bp DNA  | VRL                 |
| ACCESSION             | AF216221     | Banana bunchy top virus satellite S1 replication initiation protein (ORF VI) gene, complete cds.   | 08-MAR-2001         |
| VERSION               | AF216221.1   | GI:12004325  |                     |
| KEYWORDS              |              |  |                     |
| SOURCE                |              | Banana bunchy top virus.   |                     |
| ORGANISM              |              | Banana bunchy top virus  |                     |
| REFERENCE             |              | Viruses: ssDNA viruses; Circoviridae.  |                     |
| AUTHORS               |              | 1 (bases 1 to 1109)  |                     |
| TITLE                 |              | Horser,C., Harding,R. and Dale,J.  |                     |
|                       |              | Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein  |                     |
| JOURNAL               |              | J. Gen. Virol. 82 (Pt 2), 459-464 (2001)   |                     |
| MEDLINE               |              | 21102983   |                     |
| PUBMED                |              | 11161286   |                     |
| REFERENCE             |              | 2 (bases 1 to 1109)  |                     |
| AUTHORS               |              | Horser,C.L., Katan,M., Harding,R.M. and Dale,J.L.  |                     |
| TITLE                 |              | Additional rep-encoding DNAs associated with banana bunchy top virus   |                     |
| JOURNAL               |              | Arch. Virol. 146 (1), 71-86 (2001)   |                     |
| REFERENCE             |              | 3 (bases 1 to 1109)  |                     |
| AUTHORS               |              | Horser,C.L., Katan,M., Harding,R.M. and Dale,J.L.  |                     |
| TITLE                 |              | Direct Submission  |                     |
| JOURNAL               |              | Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia  |                     |
| FEATURES              |              |  |                     |
| source                |              | Location/Qualifiers  |                     |
|                       |              | 1..1109  |                     |
|                       |              | /organism="Banana bunchy top virus"  |                     |
|                       |              | /isolate="Taiwan"  |                     |
|                       |              | /db_xref="taxon:12585"   |                     |
|                       |              | /note="satellite S1"   |                     |
| stem_loop             |              | 3..33  |                     |
|                       |              | /note="putative sequence"  |                     |
| mRNA                  |              | 63..917  |                     |
|                       |              | /gene="ORF V1"   |                     |
| gene                  |              | 63..917  |                     |
|                       |              | /gene="ORF V1"   |                     |
| CDS                   |              | 63..917  |                     |
|                       |              | /gene="ORF V1"   |                     |
|                       |              | /codon_start=1   |                     |
|                       |              | /product="replication initiation protein"  |                     |
|                       |              | /protein_id="AAG44003.1"   |                     |
|                       |              | /db_xref="GI:12004326"   |                     |
|                       |              | /translation="MSFSRWCTLTNYSSAERDFLLALKEELNYAVGDEVAPSSQ QKHQQVLSLKSKIKLGGKKKYSRAHWRARGSDENAKYCKETLILFLGFPASQ GSHNRKLSEVNSPSPMEIEQPEIVHRYTSVKLKKFKEEFVHPCLDRPWOIQLEA IDNEPDRSLIIWYGGPNEGKSTYAKSLMKDWFTYRGKKENILFSYVDEGSEKHI VFQIPRCNODLYNDVIEALKDRVIESTKYKPKIVELINHIVMANFMPEPKCKISE DRIKIIIVC" |                     |
| BASE COUNT            | 346 a        | 213 c  | 257 g 293 t         |
| ORIGIN                |              |  |                     |
| Query Match           | 12.0%        | Score 155.2;   | DB 58; Length 1109; |
| Best Local Similarity | 51.7%;       | Pred. No. 2.6e-35;   |                     |
| Matches 461;          | Conservative | 0; Mismatches 413;   | Indels 18; Gaps 4;  |
| Qy                    | 124          | tgggtccttcaacttgaactacgacagagagagcgcggaatgttgaagcgcatc   | 183                 |
|                       |              |  |                     |
| Db                    | 78           | TGGTGCTTCACTCGAATCTACTCTCCGACGTGACGGAGAGACTTTCCTCCTCTCTG   | 137                 |
| Qy                    | 184          | gagtccttcaacttggctatgtatagttgggacgaggttgcaccttcaactggcaa   | 243                 |
|                       |              |  |                     |
| Db                    | 138          | AAGGAGGAAGAGCTTAACACGACGAGTGTCCGCGACGAGAGTCCGCGCTCCGCGCAG  | 197                 |

|                       |     |  |                   |
|-----------------------|-----|--|-------------------|
| JOURNAL               |     | Submitted (04-OCT-1993) Yu-Chan Chao, Institute of Molecular Biology, 120, Sec. 2., Yen-joe-yuan Rd., Nan-Kang, Taipei, Taiwan |                   |
| FEATURES              |     | Location/Qualifiers  |                   |
| source                |     | 1..1095  |                   |
|                       |     | /organism="Banana bunchy top virus"  |                   |
|                       |     | /specific_host="Musa acuminata Colla"  |                   |
|                       |     | /db_xref="taxon:12585"   |                   |
|                       |     | /clone="pBtd18"  |                   |
|                       |     | /note="component 2 (BBTV-C2)"  |                   |
|                       |     | 82..390  |                   |
|                       |     | /note="orf1"   |                   |
|                       |     | /codon_start=1   |                   |
|                       |     | /protein_id="AA17783.1"  |                   |
|                       |     | /db_xref="GI:475434"   |                   |
|                       |     | /translation="MSSPSLKWCFILNYSSAARENFLLKKEEDVHVAVGVDEVAPA   |                   |
|                       |     | TQKHLQGLYLSLKKMPRIEVEVWPCSLGDCQRRRRREFVLQRPNPPIRVSCC   |                   |
|                       |     | "  |                   |
|                       |     | 419..634   |                   |
|                       |     | /note="orf2"   |                   |
|                       |     | /codon_start=1   |                   |
|                       |     | /protein_id="AA17784.1"  |                   |
|                       |     | /db_xref="GI:475435"   |                   |
|                       |     | /translation="MIARSPDRMKIEQPEIFHRYOSVNLKFKKEFVHPCLDRPQWI   |                   |
|                       |     | QLTEATDEPDRSIWVYPYGNKGK"   |                   |
| BASE COUNT            |     | 337 a  | 217 c 253 g 288 t |
| ORIGIN                |     |  |                   |
| Query Match           |     | 11.1%; Score 143.6; DB 59; Length 1095;  |                   |
| Best Local Similarity |     | 49.9%; Pred. No. 8e-32;  |                   |
| Matches               |     | 451; Conservative 0; Mismatches 439; Indels 14; Gaps 3;  |                   |
| QY                    | 95  | ggctataaagggttccctccattcgccgctggtgcttcacactcgaactcagacagagag   | 154               |
| DB                    | 74  | GCCTGGAGATGTCACGTCCTCTTTAAGTGGTCTTCACTCTGAATTACTCTCCGCGG   | 133               |
| QY                    | 155 | aagagggcggaatgttgtaaggcgcatcgagtccttaacttggtctgtatgtatagttg  | 214               |
| DB                    | 134 | CAGAGAGAGAAACTTCTCTCTTCTGAAGGAGGAGGATGTTCACTACGCTGTCTCG  | 193               |
| QY                    | 215 | gggacgggttgacacttcaactggccacacacacacacacacacacacacacacac   | 274               |
| DB                    | 194 | CGCAGAGTGCCTCCGCCACCGCCAGAACACCTCCAGGGATATCTATCCCTGAAA   | 253               |
| QY                    | 275 | ccggtcgagactcaaggattgaagactgtttcttggaatgacagattcacactggagc   | 334               |
| DB                    | 254 | AGAAATGCCTCGC--GGATTGAAGAGAGATATGG--CTCCCGTGTCACTGGGAGA  | 308               |
| QY                    | 335 | cgaccctggttcgcagacagacagactactgttcgaaggaacgggtgctctcg  | 394               |
| DB                    | 309 | TTGCCAGAGGAACAGACAGAGAAATTCGAAGTACTGTTCAGAGAACCCCTAATCTCG  | 368               |
| QY                    | 395 | agcaggaagtcgcactcgctcgagtcgaaagccacgattggcccaacgatttctg  | 454               |
| DB                    | 369 | AATTAGGTTTCCCTGTTTAAATGGTCTAATAAGAGAAATATCGGAGATGATGCTC  | 428               |
| QY                    | 455 | aggaacctgatgaactccgctggaagaccacggcgagatcccggaagatggtttgacacg   | 514               |
| DB                    | 429 | GTTCTCTGATCGCATGAAATTTGAACACGCTGAGATATTTTCAGAT-----ATC   | 479               |
| QY                    | 515 | gagcttcggtggaatggacagatggccgctgaaatccggttccattccatcac  | 574               |
| DB                    | 480 | AATCTGTGAATAAGTTAAAAAATTCAGGAGGAGTTCGTTTCCTTCCTCCATAGAC  | 539               |
| QY                    | 575 | attggcacttggaagtgtctgcgtcgatcgagagcagcagcagcagcagcagcagc   | 634               |
| DB                    | 540 | CATGCCACATCAATTGACGGAGGCAATTCAGGAGAACCCGATGATCGAAGCATCATCT   | 599               |
| QY                    | 635 | ggatagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc   | 694               |
| DB                    | 600 | GGGTCTATGTCCTTATGTTAATGAGGGTAAATGAACATATCGGAAGTCAATCAAGA   | 659               |



QY 695 agccgactgtgttacacatgtgtggaaccagaaagagcgatgtaccagatcatog 754  
 Db 660 AGGATTGGTTCTACACAGGGGTGGGAAGAGAGATATCTTATCTCTATGTGGACG 719  
 QY 755 agaccacaaacgaatttaactctcgtatgtaccaggtgtaatttagatatttaatt 814  
 Db 720 AAGGATCTGACAGCATATAGTATTGATATTCCTGCTTGTATTCAGGATTATTAAT 779  
 QY 815 atgcctgttagaattgttaagacagggcattcagttcggacaaatacgaacccctta 874  
 Db 780 ATGATGTAATAGAGCATTAAGAGATAGGGTTATAGAGAGATCTAATAACAAACCATAA 839  
 QY 875 gttatcttggttcgacacatgtcgtatgtattgccaatgtcctgcctgattatt 934  
 Db 840 AGATAGTTGAATTAGTAAATATCATGTAATCGTCATGGCGAATTTTCATGCTGACTTCT 899  
 QY 935 tgaataatcagcagagacagaataaaactgtggaattatttaagatatgttcataaatt 994  
 Db 900 GTAAATCTCGAAGATCGAATAAATCAATTTATGCTGAATAAACACGCTATGACAATC 959  
 QY 995 acac 998  
 Db 960 GTAC 963

RESULT 12  
 AB000921 1009 bp DNA circular VRL 30-OCR-1998  
 LOCUS Milk vetch dwarf virus genome segment 2 encoding viral  
 DEFINITION replication-associated protein, complete sequence.  
 ACCESSION AB000921  
 VERSION AB000921.1 GI:3808180  
 KEYWORDS viral replication-associated protein.  
 SOURCE milk vetch dwarf virus (lab\_host:Pisum sativum) DNA.  
 ORGANISM Viruses; ssDNA viruses; Nanovirus.  
 Sano, Y.  
 1 (bases 1 to 1009)  
 Direct Submission  
 Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.  
 Yoshitaka Sano, Kyoto Institute of Technology, Department of  
 Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan  
 (Tel:075-724-7764, Fax:075-724-7764)  
 2 (sites)  
 Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.  
 Sequences of ten circular ssDNA components associated with the milk  
 vetch dwarf virus genome  
 J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)  
 99094638  
 On Oct 29, 1998 this sequence version replaced gi:3798653.  
 Location/Qualifiers  
 1..1009  
 /organism="milk vetch dwarf virus"  
 /viral  
 /db\_xref="taxon:67585"  
 /chromosome="segment 2"  
 /lab\_host="Pisum sativum"  
 1..32  
 69..917  
 /function="helicase"  
 /codon\_start=1  
 /product="viral replication-associated protein"  
 /protein\_id="PAA33981.1"  
 /db\_xref="GI:3798654"  
 /translation="MASKRCWCTFLNYKTLERETFLISLSDENLVFVCGDEIAPITG  
 QKHLQVSMKKLIRLGLKKFGSIHWEIAKDDOFONRKYCTKETLIAETGAPVK  
 GSNRRKIMEIYEDPEMLKRDPTDLRCKAKKRLREYCEVSVSLRPELRLRAL  
 MEVPDDRTIITWAYGPDGEGKSTFAKELIKYGFYTAGGKTQDILMTAQPDERNIAF  
 DVPRCSSEMMNYQAMEMKRNRFVASTKYRPVDLCIRKKVHLILVFANVAPDPKLSER  
 IYIING"  
 917..922  
 polyA\_signal  
 TATA\_signal 930..936  
 BASE COUNT 318 a 235 g 282 t

## ORIGIN

Query Match 10.7% Score 137.8; DB 58; Length 1009;  
 Best Local Similarity 50.3%; Pred. No. 4.4e-30;  
 Matches 430; Conservative 0; Mismatches 407; Indels 18; Gaps 3;

QY 121 cgcgtggtctcactcactgaactacgacagagagagagcgcggaatgttgaagcgc 180  
 Db 81 CGTGGTGGTCTTACGCTGAACATATAACACAGCGTTAGAGAGAGAAACATTCATTCTCTC 140  
 QY 181 atcgaagtccttaacttggtctctatagttgggaagaggttgcaacctcaactgc 240  
 Db 141 TTCTCTCGCGACGAATTTGAACATATTTCGTTTGTGGAGACGAATCGCTCCCACTACTGG 200  
 QY 241 caacgacacctccaagattccactgaacacccggtcgcgactcgcaagattgaag 300  
 Db 201 CAGAAACATCTCCAAGGATATGTTTCATGAGAAATTAATTTCGTCTGGGTGGATTGAAG 260  
 QY 301 actgtcttggaatgacaggtattccactggagccgaccccggtggttccgacgacagaa 360  
 Db 261 AAGAAATTTGGATCT---ATTGCTCATTTGGAGATCGCTAAGGGAGACGATTTTCAGAA 317  
 QY 361 agagactactgttcgaaggaacgggtgcttctcgaagacacgaggtcccgactcgtcctga 420  
 Db 318 CGCATTTACTGTACGAAAGAAACCCCTAATTGCAGAAATTTGGGCTCCCGTGAAGAAAGGC 377  
 QY 421 gtcaaaaggccacgattggcccaacgattgtcgtgaggaacctgatgaactccgctgaa 480  
 Db 378 TCAATTCGAAGGAAGATTTATGGAGATATACGAGAGAGATCCAGAGAAATGAAATTTGCGA 437  
 QY 481 gacccaggcgatcacgaagatgcgtgtcacagagcttcggtggaatggacaagatgg 540  
 Db 438 GATCCGGATACTGCCCC-----TTCGCTGTAGGGCGAAGAAATTAAGAGAACAG 485  
 QY 541 gcgcgtgaaaaatccgttccattccatatacaaatggcagcttgaagtgcgtcgtcg 600  
 Db 486 TATTGTTCTGAAGTTTCGGTGTGTTTCTTCGACCATGCAAAATTTGATTTACACAGACG 545  
 QY 601 atcggagagccagcgacgacatgcacaaatccctctgatatgcgacgagacgagagac 660  
 Db 546 TTAATGGAGGTTCCTGACGACAGAACAAATTTATTGGCGGTATGGCCCATGATGGAGCGAA 605  
 QY 661 gggaagtcggtgttggcaaatatctcgaactcaagcccgactggttctacacatgtggt 720  
 Db 606 GGTAAATCTACCTTCGCGAAGAAAT---AATCAAAATATGTTGCTTTTATCTACGAGGA 662  
 QY 721 ggaaccagaaagacgattgtaccagTaccagTaccagTaccagTaccagTaccagTaccagT 780  
 Db 663 GGAAGACCCAGGACATATCTGTATATGTATGCGCAAGATCTCTGAAGAAATATTTGCGTTC 722  
 QY 781 gatgtaccaggtgtaatttagagatttaattatgcctgtttagaattgtttaagaac 840  
 Db 723 GATGTACCCAGGTGTTCTTCGAGATGATGAATCAAGCAATGGAATGATGAAGCAAT 782  
 QY 841 agggcattcagttcggacaaatacgaaccccttagttatcttgggttcgacctatgtcat 900  
 Db 783 AGAGTATTTGCAAGTACAAATATATAGGCTCTGTAGATCTTTGTATTAAGAAAGAGGTTCA 842  
 QY 901 gtactcgtattgccaatgtcctgcctgattatttgaataatcagcagggacagaataaaa 960  
 Db 843 TTAATTTGATTGCAACGCTGGCACCTGACCCCAACAAATTAAGTAGGAGACAGAAATTGTA 902  
 QY 961 ctgtggaattattaa 975  
 Db 903 ATTATCAATTGTTGA 917

RESULT 13  
 NV5966 NV5966 1004 bp DNA circular VRL  
 LOCUS faba bean necrotic yellows virus C9 gene.  
 DEFINITION  
 ACCESSION AJ005966

```

VERSION      AJ005966.1 GI:3550532
KEYWORDS     C9 gene; component 9; putative; rep protein; replication associated
SOURCE       faba bean necrotic yellow virus.
ORGANISM     faba bean necrotic yellow virus
REFERENCE    Viruses; ssDNA viruses; Nanovirus.
REFERENCE    1 (bases 1 to 1004)
AUTHORS     Katul, L.
TITLE       Direct Submission
JOURNAL     Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und
            Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und
            Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
REFERENCE    2 (bases 1 to 1004)
AUTHORS     Katul, L., Timchenko, T., Gronenborn, B. and Vetten, H.J.
TITLE       Ten distinct circular ssDNA components, four of which encode
            putative replication-associated proteins, are associated with the
            faba bean necrotic yellow virus genome
JOURNAL     J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
MEDLINE     99094637
FEATURES     Location/Qualifiers
             1..1004
             /organism="faba bean necrotic yellow virus"
             /viral
             /isolate="SV292-88"
             /db_xref="taxon:59817"
             73..918
             /gene="C9"
             73..918
             /gene="C9"
             /function="putative replication associated (rep) protein"
             /codon_start=1
             /product="component 9"
             /protein_id="CAA06789.1"
             /db_xref="GI:3550533"
             /db_xref="SPTREMBL:O91252"
             /translation="MSAVNVVFTLNFAGEVPVLSFDERVOYAVQWHERVNHDIQCVI
            OLKKAKMNTVKNIIIGNPHLEKMGKSTIEASAYAKQESRAGVPSYGEGLKKGSHK
            RKIMELIKDPENLEPEQKYRRAMASAMDESRLKLAEEGGFPYMSWETVLGLLEE
            EPNDRIIHWYGPNGNEKSGQFGKFLGDKDLYLPGKTDQMTYMLMKNPKANVMD
            IPRCNSEYLNQFMELIKNRTIFSYKYPVGCIIINKIHVILANVLPDYKISQDRI
            KIIYC"
BASE COUNT   352 a 159 c 226 g 267 t
ORIGIN
Query Watch      10.2%; Score 132; DB 59; Length 1004;
Best Local Similarity 54.0%; Pred. No. 2.5e-28;
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
QY 476 tggagaccagcggataccgacggtgtgtacacgagcgttcggtggaatggacaa 535
DB 419 TGGAGAACCCCAAGAAATACAGAGAGCTATGCGTTGTCGCCCATGGACGAATCTCGGA 478
QY 536 gatggccgctgaaatccggtccattccatatacacaattggcagctgaagtctgt 595
DB 479 AGCTTGGCGGAAGAGAGAGCGCTTCCCTATATGTTTACAGCTGCGCAAGAAACAGTGTGG 538
QY 596 ctgcgacgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 655
DB 539 GCCTATTAGAGAGAGAGCCCAATGACCGTACTATATTGTTGGTCTACGCGCCCAATGGTA 598
QY 656 gagcgggaagtcggtgtttgccaataatctcggactcaagcccgactgggtttacacat 715
DB 599 ATGAAGGAAAAATCACAGTTGGTGTAAATCTCTGGGATTAAAGAAAGATTACCTTTATTTCAC 658
QY 716 gtggtggaaccagagagcagctatttaccatatacagacagaccacaaagaaattaa 775
DB 659 CTGAGGATAAACCCAGATATGACATATATGTTAATGAAAAATCAAGGCAAAATGTTG 718
QY 776 tccctcagtgaccagcagtgtaatttagagtatttaattatgcctctgttagaattgttta 835
DB 719 TGATGGATATCTCTGTTGTTAAATCTCAATATTAAATTAACCAATTTATGGAATTAATTA 778

```

```

QY 836 agaacaggcattcacttcggacaaatcagaaaccccttagttatcttgggttcgaccatg 895
DB 779 AAAATAGAACCATATATAGTTTAAATATGAACCACTGGATGTTATTAATAATAATAAAA 838
QY 896 tgcgtgactcgtatttgcgaatgcctcctcgtgatttggaaatcagcagggagacaa 955
DB 839 TACATGTAATTGTTATTAGCTTAATGTTATTCCTGCTGTAATGAAAAAATTAGTCAGGATAGAA 898
QY 956 taaaactgtggaatatattaa 975
DB 899 TTAATAATAATTATTGTTAA 918

RESULT 14
PBEL12187 1007 bp DNA circular VRL 16-SEP-2000
LOCUS     faba bean necrotic yellow virus C9-Eg gene, isolate Egyptian
DEFINITION EVI-93.
ACCESSION AJ132187 GI:4995171
VERSION   AJ132187.1
KEYWORDS  C9-Eg gene; rep protein.
SOURCE    faba bean necrotic yellow virus.
ORGANISM  faba bean necrotic yellow virus.
Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1007)
AUTHORS   Katul, L., Timchenko, T., Gronenborn, B. and Vetten, H.J.
TITLE     Ten distinct circular ssDNA components, four of which encode
            putative replication-associated proteins, are associated with the
            faba bean necrotic yellow virus genome
JOURNAL   J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
MEDLINE   99094637
REFERENCE 2 (bases 1 to 1007)
AUTHORS   Timchenko, T., de Kouchkovsky, F., Katul, L., David, C., Vetten, H.J.
            and Gronenborn, B.
TITLE     A single rep protein initiates replication of multiple genome
            components of faba bean necrotic yellow virus, a single-stranded
            DNA virus of plants
JOURNAL   J. Virol. 73 (12), 10173-10182 (1999)
MEDLINE   20027244
REFERENCE 3 (bases 1 to 1007)
AUTHORS   Katul, L.
TITLE     Direct Submission
JOURNAL   Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie,
            Mikrobiol., u. biol. Sicherheit, Biologische Bundesanstalt f. Land-
            u. Forstwirtschaft, Messeweg 11 -12, 38104 Braunschweig, GERMANY
FEATURES  Location/Qualifiers
             1..1007
             /organism="faba bean necrotic yellow virus"
             /viral
             /isolate="Egyptian EVI-93"
             /db_xref="taxon:59817"
             /lab_host="Vicia faba"
             1..39
             73..1003
             /gene="C9-Eg"
             73..918
             /gene="C9-Eg"
             /function="replication-associated protein"
             /note="32.53 kDa"
             /codon_start=1
             /evidence="experimental"
             /product="Rep protein"
             /protein_id="CAB44027.1"
             /db_xref="GI:4995172"
             /db_xref="SPTREMBL:O9WIK2"
             /translation="MSAVNVVFTLNFAGEVPVLSFDERVOYAVQWHERVNHDIQCVI
            OLKKAKMNTVKNIIIGNPHLEKMGKSTIEASAYAKQESRAGVPSYGEGLKKGSHK
            RKIMELIKDPENLEPEQKYRRAMASAMDESRLKLAEEGGFPYTLWSQETVLGLLEE
            EPNDRIIHWYGPNGNEKSGQFGKFLGDKDLYLPGKTDQMTYMLMKNPKANVMD
            IPRCNSEYLNQFMELIKNRTIFSYKYPVGCIIINKIHVILANVLPDYKISQDRI
            KIIYC"
             939..944
             /gene="C9-Eg"
TATA_signal

```

```

polyA_signal 998..1003
BASE COUNT 353 a 161 c 227 g 266 t
ORIGIN

Query Match 10.2%; Score 132; DB 58; Length 1007;
Best Local Similarity 54.0%; Pred. No. 2.5e-28;
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 476 tgaagaccgagcggtacacgaagatcggtgtacacgagagctcggtggaatggacaa 535
Db 419 TGGAGAACCCAGAAATACAGAGAGCGATGGCTGTGCGCCATGACGACCAATCTCGGA 478
QY 536 gatggcgctgaaatccgttccattccatatacacaattggcagcttgaaagtctgt 595
Db 479 AGCTTGCTGAAGAGAGGCGCTTCCCTATACGCTTTACAGCTGGCAAGAACAGTGTGG 538
QY 596 ctgcgatcggagacgacgacgacgacgacgacgacgacgacgacgacgacgacgag 655
Db 539 GCCTATTAGAAGAGAGAGCCCAATGACCGTATTATTATTGGTCTACGGCCCAATGGTA 598
QY 656 gagcgggaagtcggtgttccaaatatctcggaataacgacgacgacgacgacgacgac 715
Db 599 ACGAAGGAAATACAGATTGCTAAATTCCTGGGATTAAATAAAGATTTACCTTTATTAC 658
QY 716 gtgtggacacgaagagcgtattgtaccagtagacgacgacgacgacgacgacgacgaa 775
Db 659 CTGGAGGTAAACCCCAAGATATGATATATGTTAATGAAATCCCAAGGCAATGTG 718
QY 776 tctcgtatgaccaggtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 835
Db 719 TGATGGATATTCCTCGTGTGTAATTCGAAATTTAAATTTATCAATTTATGAATTAATA 778
QY 836 agaacaggcattcagttcgcgaacaaatcagaccccttagttatcttgggttgcacatg 895
Db 779 AAATAGAACCAATATTAGTTTAAATATGAAACAGTGGATGATTTAATAAATAAATA 838
QY 896 tgcgtactcgtattgccaatgcctgcctgattattgaaatcagcagcagcagcagcag 955
Db 839 TACATGTAATTTAGTAAATGATTTGCTGATTTGCTGATTTGCTGATTTGCTGATTTGCT 918
QY 956 taaactgtggaatttaa 975
Db 899 TTAATAATTTATTTGTTAA 918

RESULT 15
LOCUS NVV5968
DEFINITION faba bean necrotic yellows virus CI-Eg gene. 13-JAN-1999
ACCESSION AJ005968
VERSION AJ005968.1 GI:3550536
KEYWORDS CI-Eg gene; component 1-Eg; putative; rep protein; replication associated protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1004)
AUTHORS Katul,L.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
REFERENCE 2 (bases 1 to 1004)
AUTHORS Katul,L., Timchenko,T., Gronenborn,B. and Vetter,H.J.
TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
MEDLINE 99094637
FEATURES
    Location/Qualifiers
        source
            1..1004

```

```

/organism="faba bean necrotic yellows virus"
/virion
/isolate="EVI-93"
/db_xref="taxon:59817"
1..31
stem_loop
63..1003
gene
63..911
CDS
/genes="CI-Eg"
/function="putative replication associated (rep) protein"
/codon_start=1
/product="component 1-Eg"
/protein_id="CAA06791.1"
/db_xref="GI:3550537"
/db_xref="SPTREMBL:O91254"
/translation="MASKRCFTLNYKTAVERESFISLFSRDELNYFVGDETAFTPTN
OHKQYVSLKMIKRLGILKKFGYRAHWEIAKRGDDQFNDRDYCTKETLISEIGAPVK
GSKQRKIMDLVQDPEMOLKDPDTALRCNAKRLRIEYCSSFAVLSLWSELRHVL
MAEPDDRTIIWVYSGDGRGKSTFAKELIKYWEYTAGKTDQILLYMAQDPERNIAP
DVPRCSSEMMNYQAMEMLNKRVFASTKTRPVLCVRKVLHLLVFANVSPDPKISEDR
IVINC"
polyA_signal
911..916
/genes="CI-Eg"
TATA_signal
998..1003
/gene="CI-Eg"
BASE COUNT 305 a 179 c 224 g 296 t
ORIGIN

Query Match 9.5%; Score 122.8; DB 59; Length 1004;
Best Local Similarity 49.2%; Pred. No. 1.4e-25;
Matches 421; Conservative 0; Mismatches 417; Indels 18; Gaps 3;

QY 121 cgtcgttctcactcgaactacgagagagagagagagagagagagagagagagagagc 180
Db 75 CGCTGCTGCTTACTTTGAACATAAAGAGCTGTAGAGAGAGAGAGAGAGAGAGAGAGAG 134
QY 181 atcagagtccttaacttgctcgtatagttgggagcagagagagagagagagagagag 240
Db 135 TTTCCTCGAGAGCAATTAACATATTTGCTGCGGAGAGCAAACTGCTCTACTACGAAT 194
QY 241 caacgacaccccaagagattccacttgaagagcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
Db 195 CAAAACATCTCAAGGTTATGTTTCGTTGAAGAAATGATTCGCTCGGTGGATTGAA 254
QY 301 actgttcttgggaatgacagagattcacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
Db 255 AAGAAATTTGG--TTATCGTCTCATTTGGAAATTTGCCAAAGAGAGAGAGAGAGAG 311
QY 361 agagactactgtcgaagagagagagagagagagagagagagagagagagagagagagag 420
Db 312 CGCGACTACTGTACCAAGAAACCCCTAATTTCTGAATTTGGGGCTCCTGTGAAGAAAGGC 371
QY 421 gtcaaaagccacgattggcccaacgattgtcgtgaggaacacgtgagacacacacacac 480
Db 372 TCGAATCAAAAGAAAGATTATGGATTATATCTTCAAGATCCCGAAGAAATGCAATTGAAG 431
QY 481 gacccagcggataccgaagatgcgtttacacgagagcttcggtggaatgagacagagag 540
Db 432 GATCCAGATACATGCTCTTCGCTGCTAACCGAAGAGATTGAGAAATTTGATTTGTCGTCG 491
QY 541 gccgtgaaaaatccgttccattccattccattccattccattccattccattccattcc 600
Db 492 TTCGCTGTGATATCCCT-----TCGGCCATGGCAGCTCTGACGCTTCATCGGGTT 539
QY 601 atcggagagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 660
Db 540 TTAATGGCAGAACACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 599
QY 661 ggaagagtcggtgttggccaaatatctcggactcaagccgactggttctacacatgtggt 720
Db 600 GGTAATAACCTTCGCGAAGGAAT---AATTAAGTATGATGTTTATTAATCTACTGCAGGA 656

```

QY 721 ggaaccagaaggaagcgattgtaccagttacatcgaggaccacaaacgaaatttaactc 780  
 Db 657 GGAAGAGCGCAGGATATTCTGTATATGTATGCTCAAGACCCAGAGAGAAATATCGCGTTC 716  
 QY 781 gatgtaccaggtgtaatttagagttatttaattatgcccctgttagaattgtgttaagaac 840  
 Db 717 GATGTTCACAGGTGTTCTTCGAGATGATGAACATATCAAGCTATGGAGATGTTGAAGAAC 776  
 QY 841 agggcattcagttcggacaatacgaaccccttagttatcttgggttcgaccatgtgcat 900  
 Db 777 AGAGTATTGCAAGTACAAAATATAGGCCCGTAGATCTTTGTGTAGAAAGAAAGTTCAI 836  
 QY 901 gtactcgtatttgccaatgctcctgctgattatttgaaatcagcaggacagataaaa 960  
 Db 837 TTAATTGTTGTGCAACGTCACCTGACCCACAAAAATAAGTGAGGACAGAAATTGTA 896  
 QY 961 ctgtggaatatttaa 976  
 Db 897 ATTATCAATTGTTGAA 912

Search completed: October 24, 2001, 11:42:08  
 Job time: 6313 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:39 ; Search time 5701.85 Seconds  
(without alignments)  
2118.665 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_211\_991  
Perfect score: 781  
Sequence: 1 gttgggacgaggttgcacc.....ttaaagtatgtcatctaa 781

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_om.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_pil.\*
- 13: gb\_pil2.\*
- 14: gb\_pil3.\*
- 15: gb\_pil4.\*
- 16: em\_ba1.\*
- 17: em\_ba2.\*
- 18: em\_fun.\*
- 19: em\_htgo\_hum.\*
- 20: em\_htgo\_inv.\*
- 21: em\_htgo\_rod.\*
- 22: em\_htg\_hum1.\*
- 23: em\_htg\_hum2.\*
- 24: em\_htg\_hum3.\*
- 25: em\_htg\_hum4.\*
- 26: em\_htg\_hum5.\*
- 27: em\_htg\_hum6.\*
- 28: em\_htg\_hum7.\*
- 29: em\_htg\_hum8.\*
- 30: em\_htg\_inv1.\*
- 31: em\_htg\_inv2.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_rod.\*
- 34: em\_hum1.\*
- 35: em\_hum2.\*
- 36: em\_hum3.\*
- 37: em\_hum4.\*
- 38: em\_hum5.\*
- 39: em\_hum6.\*
- 40: em\_hum7.\*
- 41: em\_in.\*
- 42: em\_om.\*
- 43: em\_or.\*

- 44: em\_ov.\*
- 45: em\_pat.\*
- 46: em\_ph.\*
- 47: em\_pil.\*
- 48: em\_ro.\*
- 49: em\_sts.\*
- 50: em\_sy.\*
- 51: em\_un.\*
- 52: em\_vi.\*
- 53: gb\_sts1.\*
- 54: gb\_sts2.\*
- 55: gb\_sts3.\*
- 56: gb\_sy.\*
- 57: gb\_un.\*
- 58: gb\_vil.\*
- 59: gb\_vil2.\*
- 60: gb\_vil3.\*
- 61: gb\_vil4.\*
- 62: gb\_vil5.\*
- 63: gb\_vil6.\*
- 64: gb\_vil7.\*
- 65: gb\_vil8.\*
- 66: gb\_vil9.\*
- 67: gb\_vil10.\*
- 68: gb\_vil11.\*
- 69: gb\_vil12.\*
- 70: gb\_vil13.\*
- 71: gb\_vil14.\*
- 72: gb\_vil15.\*
- 73: gb\_vil16.\*
- 74: gb\_vil17.\*
- 75: gb\_vil18.\*
- 76: gb\_vil19.\*
- 77: gb\_vil20.\*
- 78: gb\_vil21.\*
- 79: gb\_vil22.\*
- 80: gb\_vil23.\*
- 81: gb\_vil24.\*
- 82: gb\_vil25.\*
- 83: gb\_vil26.\*
- 84: gb\_vil27.\*
- 85: gb\_vil28.\*
- 86: gb\_vil29.\*
- 87: gb\_vil30.\*
- 88: gb\_vil31.\*
- 89: gb\_vil32.\*
- 90: gb\_vil33.\*
- 91: gb\_vil34.\*
- 92: gb\_vil35.\*
- 93: gb\_vil36.\*
- 94: gb\_vil37.\*
- 95: gb\_vil38.\*
- 96: gb\_vil39.\*
- 97: gb\_vil40.\*
- 98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------|-------------|
| 1          | 781   | 100.0   | 1291         | 58    | CFDCG       |
| 2          | 162.6 | 20.8    | 1096         | 9     | AR063452    |
| 3          | 162.6 | 20.8    | 1096         | 58    | BYTV2       |
| 4          | 156   | 20.0    | 1095         | 58    | AF216222    |
| 5          | 142.6 | 18.3    | 1091         | 9     | AR063453    |
| 6          | 142.6 | 18.3    | 1111         | 58    | BB012586    |
| 7          | 141   | 18.1    | 1127         | 58    | BB012587    |
| 8          | 137.8 | 17.6    | 1106         | 9     | AR063451    |

M29963 Coconit fol  
AR063452 Sequence  
L32167 Banana bunc  
AF216222 Banana bu  
AR063453 Sequence  
U12586 Banana bunc  
U12587 Banana bunc  
AR063451 Sequence

|    |       |      |      |    |            |          |             |
|----|-------|------|------|----|------------|----------|-------------|
| 9  | 137.8 | 17.6 | 1106 | 58 | BYTV1      | L32166   | Banana bunc |
| 10 | 136.2 | 17.4 | 1109 | 58 | AF216221   | AF216221 | Banana bu   |
| 11 | 132   | 16.9 | 1004 | 59 | NVV5966    | AJ005966 | faba bean   |
| 12 | 132   | 16.9 | 1007 | 58 | FBE132187  | AJ132187 | faba bean   |
| 13 | 119.6 | 15.3 | 1095 | 59 | U02312     | U02312   | Banana bunc |
| 14 | 114.6 | 14.7 | 1009 | 58 | AB000921   | AB000921 | Milk vetc   |
| 15 | 111.2 | 14.2 | 1022 | 58 | SCU16731   | U16731   | Subterranea |
| 16 | 104   | 13.3 | 1015 | 58 | FBE132185  | AJ132185 | faba bean   |
| 17 | 104   | 13.3 | 1022 | 58 | AB009047   | AB009047 | Milk vetc   |
| 18 | 103.4 | 13.2 | 1004 | 59 | NVV5968    | AJ005968 | faba bean   |
| 19 | 101.6 | 13.0 | 1014 | 59 | NVV5964    | AJ005964 | faba bean   |
| 20 | 99.2  | 12.7 | 1017 | 59 | SCU16735   | U16735   | Subterranea |
| 21 | 96.4  | 12.3 | 1002 | 58 | FBNY9RE2   | X80879   | FBNVR gene  |
| 22 | 96    | 12.3 | 1000 | 58 | AB000922   | AB000922 | Milk vetc   |
| 23 | 92.8  | 11.9 | 1007 | 58 | AB000920   | AB000920 | Milk vetc   |
| 24 | 91.4  | 11.7 | 1375 | 58 | CJEF132345 | AJ132345 | Nanovirus   |
| 25 | 88.2  | 11.3 | 1367 | 58 | AYE238493  | AJ238493 | Nanovirus   |
| 26 | 86.6  | 11.1 | 1376 | 58 | CJEF132344 | AJ132344 | Nanovirus   |
| 27 | 82.8  | 8.0  | 1110 | 9  | AR010230   | AR010230 | Sequence    |
| 28 | 59.6  | 7.6  | 982  | 9  | AR112007   | AR112007 | Sequence    |
| 29 | 59.6  | 7.6  | 1110 | 9  | AR010227   | AR010227 | Sequence    |
| 30 | 59.6  | 7.6  | 1110 | 58 | BBU18077   | U18077   | Banana bunc |
| 31 | 59.6  | 7.6  | 1111 | 9  | AR010225   | AR010225 | Sequence    |
| 32 | 59.6  | 7.6  | 1111 | 9  | AR010228   | AR010228 | Sequence    |
| 33 | 59.6  | 7.6  | 1111 | 9  | AR010240   | AR010240 | Sequence    |
| 34 | 59.6  | 7.6  | 1111 | 58 | AF102780   | AF102780 | Banana bu   |
| 35 | 59.6  | 7.6  | 1111 | 59 | S56276     | S56276   | orf v1 {com |
| 36 | 58.6  | 7.5  | 1103 | 9  | AR010232   | AR010232 | Sequence    |
| 37 | 58.6  | 7.5  | 1105 | 9  | AR010234   | AR010234 | Sequence    |
| 38 | 58    | 7.4  | 1109 | 9  | AR010229   | AR010229 | Sequence    |
| 39 | 58    | 7.4  | 1111 | 9  | AR010231   | AR010231 | Sequence    |
| 40 | 57    | 7.3  | 1103 | 58 | AF238874   | AF238874 | Banana bu   |
| 41 | 57    | 7.3  | 1103 | 59 | U97525     | U97525   | Banana bunc |
| 42 | 57    | 7.3  | 1104 | 9  | AR010233   | AR010233 | Sequence    |
| 43 | 56.4  | 7.2  | 1111 | 9  | AR010226   | AR010226 | Sequence    |
| 44 | 53.8  | 6.9  | 1103 | 58 | AF238875   | AF238875 | Banana bu   |
| 45 | 52.2  | 6.7  | 1103 | 58 | AF246123   | AF246123 | Banana bu   |

```

RESULT 1
  CFCGC      1291 bp ss-DNA circular VRL
  Coconut foliar decay virus, complete genome.
  M29963
  M29963.1 GI:323306
  circular; complete genome.
  Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
  Coconut foliar decay virus
  Viruses; ssDNA viruses; Nanovirus.
  1 (bases 1 to 1291)
  Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
  Nucleotide sequence of a circular single-stranded DNA associated
  with coconut foliar decay virus
  Virology 176, 648-651 (1990)
  90266484
  Draft entry and printed sequence for [1] kindly submitted by
  W.Rhode, 15-MAR-1989, for release after publication.
  Location/Qualifiers
    1..1291
      /organism="Coconut foliar decay virus"
      /db_xref="taxon:12474"
      40..70
        /note="stem-loop structure"
      103..975
        /note="ORF 1"
      /codon_start=1
      /protein_id="AAA42894.1"
      /db_xref="GI:323307"
      /translation="MGSSIRKWCFTLNYETEEEAANVVRRIESLNLYAIVGDENVAPS
      TCGPHLOCFIHLKTCPRFQIKLTVIGNDRILFETRGSDQNRDYGCKSERVILLEGHPV

```

| Query Match           | 100.0% | Score  | 781;          | DB         | 58; | Length | 1291; |
|-----------------------|--------|--|---------------|------------|-----|--------|-------|
| Best Local Similarity | 100.0% | Pred.  | No. 2.5e-221; |            |     |        |       |
| Matches               | 781;   | Conservative   | 0;            | Mismatches | 0;  | Indels | Gaps  |
| QY                    | 1      | gttgggacgaagtgcaccttcaactgccacgacacactccaagattcatccacttg       | 60            |            |     |        |       |
| Db                    | 211    | ctttggggacgaggttgcacctttcaacttggccacacgacctccaaggattcatccacttg | 270           |            |     |        |       |
| QY                    | 61     | aagaccggtcggcgactgcgaaggttgaaactgtttcttgggaatgacaggattcaccttg  | 120           |            |     |        |       |
| Db                    | 271    | agaccggtcggcgactgcgaaggttgcaagactgtttcttgggaatgacaggattcaccttg | 330           |            |     |        |       |
| QY                    | 121    | gagccgacccctggtttccgacgaacaaatagagactactgttcgaagcaacgggtgctt   | 180           |            |     |        |       |
| Db                    | 331    | gagccgacccctggtttccgacgaacaaatagagactactgttcgaagcaacgggtgctt   | 390           |            |     |        |       |
| QY                    | 181    | ctcgaagcaggaagtcgccactcgtctctggagtcaaaaggccacgattggcccaacgattt | 240           |            |     |        |       |
| Db                    | 391    | ctcgaagcaggaagtcgccactcgtctctggagtcaaaaggccacgattggcccaacgattt | 450           |            |     |        |       |
| QY                    | 241    | gctggaagacctgtatgaactccgctctggaagaccacggcgagataccgaagatcgcttga | 300           |            |     |        |       |
| Db                    | 451    | gctggaagacctgtatgaactccgctctggaagaccacggcgagataccgaagatcgcttga | 510           |            |     |        |       |
| QY                    | 301    | cacgaagctcggatggaatgacaagaatggcgcgctgaaatccggttcccatctccatat   | 360           |            |     |        |       |
| Db                    | 511    | cacggagcttccggtggatggacagatggcgcgctgaaatccggttcccatctccatat    | 570           |            |     |        |       |
| QY                    | 361    | cacaattggcagcttgaagtgtctctgcgactcgagagaccgagcgagcgtcgcaaatc    | 420           |            |     |        |       |



QY 481 ctcaagcccgactggttctacacatgtggtggaaccagaaggacgtattgtaccagtac 540

ORIGIN

| EDRKLING   |       |       |
|------------|-------|-------|
| BASE COUNT | 347 a | 219 c |
| ORIGIN     |       |       |
|            | 252 g | 277 t |





|             |  |  |                              |
|-------------|--|--|------------------------------|
| Ddb         | 470  | -ACATCTGTGAAGAAGTTAAAAAAATTCGAAGGAGGAATTCGTTCATCTCTGCCTCGAT  | 528                          |
| Qy          | 361  | cacaattgcagctgaagtctgtctcgatcggagagccagcggagcgtacgcacaatc    | 420                          |
| Ddb         | 529  | AGACCATGCGCATTCATTTGACGGAGGCAATTGACGAGGAACAGATGATCGAAGCATC   | 588                          |
| Qy          | 421  | ctctggaatcggcagcagcggagagcagcgggaagtcgctgtgttgccaaatatctcgga | 480                          |
| Ddb         | 589  | TTCTGGGTCTATGTTCCGAATGGTAAATGAGGGGAATCAACATATGCGAAGTCATT---  | 645                          |
| Qy          | 481  | ctcaagcccgactggtttacacactgagtggaaaccgaaagcagctattgtaccagtac  | 540                          |
| Ddb         | 646  | ATCAAGAAGGACTGGTTCTACACGAGAGTGGGAGGAGGAGGAGACATACTGTCTCTTAC  | 705                          |
| Qy          | 541  | atcgagcagcccaaaacgaa--atttaactcctcgatgtaccacagtgtaattgaagtat | 597                          |
| Ddb         | 706  | GTGGCAGAGGATCTGAGAAGCATATTGATTGTATATTCCTCGTGTAAATCAGATTAT    | 765                          |
| Qy          | 598  | ttaaatatgcctgttgaagtgttgaagacacagcattcagtcgagacaaatcacaa     | 657                          |
| Ddb         | 766  | TTAAATATGATGTTATAGAGGCATTAAAGGATAGGGTCATAGAGAGTACTAAATATAA   | 825                          |
| Qy          | 658  | cccttagttatcttgggttcgacctgtgcattgtactgtatttgccaatgctcgtgct   | 717                          |
| Ddb         | 826  | CCTATTAACTTACTTGAATTAATATATACATGTAATTTGTAATGCTAAATTCATGCOA   | 885                          |
| Qy          | 718  | gattattgaaatacagcgggacagaataaaaactgtggaattataa               | 766                          |
| Ddb         | 886  | GAATTCTGTAATAATCTCCGAGATAGAAATTAAGATTATTATTGTTAAA            | 934                          |
| RESULT      | 7  |  |                              |
| BBU12587    |  |  |                              |
| LOCUS       | BBU12587   | 1127 bp  | DNA circular VRL 01-FEB-1995 |
| DEFINITION  | Banana bunchy top virus  | DNA IV ORF VI and ORF CI genes, complete cds.                |                              |
| ACCESSION   | U12587   | GI:642393  |                              |
| VERSION     | U12587.1   |  |                              |
| KEYWORDS    | Banana bunchy top virus.   |  |                              |
| SOURCE      | Banana bunchy top virus.   |  |                              |
| ORGANISM    | Banana bunchy top virus.   |  |                              |
| REFERENCE   | Viruses; ssDNA viruses; Circoviridae.  |  |                              |
| AUTHORS     | Wu, R.-Y. and You, L.-R.   |  |                              |
| TITLE       | Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related virus DNAs  |  |                              |
| JOURNAL     | Unpublished  |  |                              |
| REFERENCE   | 2 (bases 1 to 1127)  |  |                              |
| AUTHORS     | Wu, R.-Y.  |  |                              |
| TITLE       | Direct Submission  |  |                              |
| JOURNAL     | Submitted (25-JUL-1994) Rey-Yuh Wu, Agricultural Biotechnology Division, Development Center for Biotechnology, 81 Chang Hsing Street, Taipei, 10671, Taiwan, ROC   |  |                              |
| FEATURES    | Location/Qualifiers  |  |                              |
| source      | 1..1127  |  |                              |
|             | /organism="Banana bunchy top virus"  |  |                              |
|             | /db_xref="taxon:12585"   |  |                              |
|             | /clone="S15 and 4-2"   |  |                              |
| TATA_signal | 1..8   |  |                              |
| stem_loop   | 15..49   |  |                              |
| CDS         | 79..933  |  |                              |
|             | /note="ORF V1"   |  |                              |
|             | /codon_start=1   |  |                              |
|             | /protein_id="AAA61877.1"   |  |                              |
|             | /db_xref="GI:642394"   |  |                              |
|             | translation="MSGFKWCFTLNSYSAERDFLLAKKEELNYAVGVGEVAPSSGQKHLQYLKLSIKIGLKKKYSSRAHWERARCGSDENAKYCSKETLILELGPASQKSNRKLSEWVSFPERMRIRPEIYHYTSVKLKKFKEFVHPCLDROPWQIOLTEATVDEPDDRTIFWYTPNGNEGKSYIAKSLMKDWFTYEGCKKNILPSYDGESEKHIIFDPICRNQDLYNDVIEKADRVIESKYKPIKLVELNLNHIIVMANFPEFCKIEEDRIKIYC" |  |                              |
|             | complement(497..643)   |  |                              |



[illegible]





```

LOCUS       U02312                1095 bp    DNA                    VRL          28-APR-1994
DEFINITION  Banana bunchy top virus component 2.
ACCESSION   U02312
VERSION     U02312.1   GI:475433
KEYWORDS    .
SOURCE      Banana bunchy top virus.
ORGANISM   Banana bunchy top virus.
            Viruses; ssDNA viruses; Circoviridae.
REFERENCE   1  (bases 1 to 1095)
AUTHORS    Yeh,H.H., Su,H.J. and Chao,Y.
TITLE      Genome characterization and identification of viral-associated
            dsDNA component of banana bunchy top virus
JOURNAL     Virology 198, 645-652 (1994)
MEDLINE     94120739
REFERENCE   2  (bases 1 to 1095)
AUTHORS    Chao,Y.
TITLE      Direct Submission
JOURNAL     Submitted (04-OCT-1993) Yu-Chan Chao, Institute of Molecular
            Biology, 120, Sec. 2., Yen-Joe-yuan Rd., Nan-Kang, Taipei, Taiwan
            115, Republic of China
FEATURES             Location/Qualifiers
     1..1095
         /organism="Banana bunchy top virus"
         /specific_host="Musa acuminata Colla"
         /db_xref="taxon:12585"
         /clone="pBTD18"
         /note="component 2 (BETV-C2)"
     82..1390
         /note="orf1"
         /codon_start=1
         /protein_id="AA17783.1"
         /db_xref="GI:475434"
         /translation="WSSPSLKWCFITLYSSAARENFLSLKKEEDVHYAVVGDEVAPA
            TQGHLOGVLISLKKMPRIEEVWLPCLGDCORNRREFVLFORNPNRIVRSC
            419..634
            /note="orf2"
            /codon_start=1
            /protein_id="AA17784.1"
            /db_xref="GI:475435"
            /translation="MIARSPDRMKTEQPEIFHRYQSVNKLKKEEFVHPCIDLRPWQI
            QUTEAIDEEPDRSIIVYGYGNEGK"
     337 a      217 c   253 g   286 t
BASE COUNT
ORIGIN

```

[illegible]

|            |   |  |                              |
|------------|---|--|------------------------------|
| Qy         | 361   | cacaattggcagctgaagtgtctgcgatacgagacgagcaggacgatcgacaaatc       | 420                          |
| Db         | 536   | AGACCATGGCAGATACAATTGACGGAGGCAATTGACGAGGAACTCCGATGATCGAAGCATC  | 595                          |
| Qy         | 421   | ctctgatatcgagcagacgagacgagacgaggaagtcctgttggccaaatatctcgga     | 480                          |
| Db         | 596   | ATCTGGGTCTATGGTCTTATGGTAATGAGGCTAAATGAACATATGCCAAGTCATAATC     | 655                          |
| Qy         | 481   | ctcaagcccgaactggtttcacatcttggtggaaccgagaagagcgttatgtaccagtac   | 540                          |
| Db         | 656   | AAGAAGGATTTGTTCTACACAGGGTGCGAAGAGAGGAAATATCTTTCTCTCTATGTG      | 715                          |
| Qy         | 541   | atcgagagccccaaaacgaaattaatctctcatgatlaccagggtgtaatttagagtattta | 600                          |
| Db         | 716   | GACGAAGGATCTGCACAAGCATATAGTATTTCATCTCGTTGTAATCAGGATTATTTA      | 775                          |
| Qy         | 601   | aattatgccctgttagaagtgttaaacaaggcgcattcagttcggaacaatacgaacctc   | 660                          |
| Db         | 776   | AATTATGATGTAAATAGAGGCCATTAAAGCATAGGGTTTAGAGACTACTAAATACAACC    | 835                          |
| Qy         | 661   | cttaagtctcttggttgcgacctgtgcattctactctatttccaatgctcctgcctgat    | 720                          |
| Db         | 836   | ATAAAGATAGTTGAATTAGGTAAAATACATCTAATCTCATGCGAATTCATGCGCTGAC     | 895                          |
| Qy         | 721   | tatttgaataatcagcagggacagaataaaaaact                            | 752                          |
| Db         | 896   | TTCGTAAATCTCGAAGATCGAATAAAAAAT                                 | 927                          |
| <br>       |   |  |                              |
| RESULT     | 14  |  |                              |
| AB0000921  |   |  |                              |
| LOCUS      | AB0000921   | 1009 bp  | DNA circular VRL 30-OCT-1998 |
| DEFINITION | Milk vetch dwarf virus genome segment 2 encoding viral replicase-associated protein, complete sequence.   |  |                              |
| ACCESSION  | AB0000921   |  |                              |
| VERSION    | AB0000921.1   | GI:3808180   |                              |
| KEYWORDS   | viral replication-associated protein.   |  |                              |
| SOURCE     | milk vetch dwarf virus (lab_host:Pisum sativum) DNA.  |  |                              |
| ORGANISM   | Viruses: ssDNA viruses; Nanovirus.  |  |                              |
| REFERENCE  | 1 (bases 1 to 1009)   |  |                              |
| AUTHORS    | Sano,Y.   |  |                              |
| TITLE      | Direct Submission   |  |                              |
| JOURNAL    | Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.   |  |                              |
| REFERENCE  | Yoshitaka Sano, Kyoto Institute of Technology, Department of Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan (Tel:075-724-7764, Fax:075-724-7764) |  |                              |
|            | 2 (sites)   |  |                              |
|            | Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M. Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome         |  |                              |
| AUTHORS    | J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)   |  |                              |
| TITLE      | 99094638  |  |                              |
| JOURNAL    | On Oct 29, 1998 this sequence version replaced gi:3798653.  |  |                              |
| MEDLINE    | location/Qualifiers   |  |                              |
| COMMENT    | 1..1009   |  |                              |
| FEATURES   | /organism="milk vetch dwarf virus"  |  |                              |
| source     | /virion   |  |                              |
|            | /db_xref="taxon:67585"  |  |                              |
|            | /chromosome="segment 2"   |  |                              |
|            | /lab_host="Pisum sativum"   |  |                              |
|            | 1..32   |  |                              |
|            | 69..917   |  |                              |
|            | /function="helicase"  |  |                              |
| stem_loop  |   |  |                              |
| CDS        |   |  |                              |

```

/product="viral replication-associated protein"
/protein_id="PAA33961.1"
/db_xref="GI:3798654"
/translation="MASKRKLTLKALERTFLSRDELNYFVGCGDEIAPTTG
KQHLOGVSKMLRLGLGKKKGTSAHWEIAKDDQNRDYCTKTLIAETGAPVK
GNNRKLMEIYEDPEMKLRDPDTRLCRAKKLRREYSEVSLRSPWQELHRL
MEVPDDRTIWIAYDGGEGCKTFAKELIKYGFYAGTKTDILYMAADPERNIAP

```

Thu Oct 25 13:07:56 2001

DVPRCSSEMMNYQAMEMKNNRVFASTKYRPVLDLIRKKVHLVIFANVAPDPTKLSDR  
IIVINC

polyA\_signal 917..922  
TATA\_signal 930..936

BASE COUNT 318 a 174 c 235 g 282 t  
ORIGIN

Query Match 14.7%; Score 114.6; DB 58; Length 1009;  
Best Local Similarity 49.9%; Pred. No. 5.3e-23;  
Matches 381; Conservative 0; Mismatches 364; Indels 18; Gaps 3;

QY 3 tgggacaggtgtgacactcaactgcccacacacacactcccaaggattccacactgaa 62  
Db 173 TGGAGCAAGAAATCGTCCACTACTGGGAGAAACATCTCAAGGATATGTTTCGATGAA 232

QY 63 gaccggtcggaactgcaagattgaagactgtcttgggaatgacagagattcacctgga 122  
Db 233 GAAATTAATTCGTGGTGGATTGAAGAGAAATTTGGATCT---ATTGCTCATTTGGGA 289

QY 123 gccgaccgtgttcgcagcaacagaaatagagactactgttcgaagaaacgggtgtctt 182  
Db 290 GATCGTAGAGGAGAGGATTTTCAAGATTCGGATTACTGTACGAGAAAGAAACCCCTAATGC 349

QY 183 cgagcagcaggtcccgactcgtcctggtgagtcacaaaggccacgattggcccaacgattgc 242  
Db 350 AGAAATTTGGGCTCCGGTGAAGAAAGGCTCAAAATCGAAGAGATTTGAGGATATACGA 409

QY 243 tgaggaaactgatgaactccgctggaagaccagcggcggaataccgaagatcggtgtaca 302  
Db 410 AGAAGATTCAGAGAAATGAATTTGGAGATCCGATGACTGCCCTTCGCTGAAG--- 465

QY 303 cggagctcgtggaatgacgaagatggccgctgaaatccgttcccaattccatata 362  
Db 466 -----CGAAGAAATTAAGAGAGATTTGTTCTGAAAGTTTCGGGTGTTTCTCTTCG 517

QY 363 caattggcagctgaagtgtctgtcgaatcgagcagcggcgagatcgcaacaactc 422  
Db 518 ACCATGCAAAATTTGAGTTACAGACGCTTAATGAGGTTCTTGACGACAGAAATAT 577

QY 423 ctggatgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 482  
Db 578 TTGGCGGTATGGCCAGATGGAGCGCAAGTAAATCTACTTTCGCGAAAGAAAT---AAT 634

QY 483 caagccgactgtttacacatgtgtggaaccagacgagcagcagcagcagcagcagcagcagc 542  
Db 635 CAAATATGTTGTTTATCTACGAGGAGAAAGACCCAGGACATCTGTATATGTATGC 694

QY 543 cgaggacccaaacgaatttaactcctcgtatgaccagcagcagcagcagcagcagcagcagcagc 602  
Db 695 GCAAGATCTGAAAGAAATATTTGGTTCGATGTACCCAGGTTCTTCGGAGATGATGAA 754

QY 603 ttatgacctgtgaagtgttaagaacagcagcagcagcagcagcagcagcagcagcagcagcagc 662  
Db 755 CTATCAAGCAATGAAGATGATGAAGATGAGATTTTGAAGTACAAATATAGGCTGT 814

QY 663 tagttatcttgggttcgacctgtgcatctactcgtatgttggcaatgtcctgcctgatta 722  
Db 815 AGATCTTTGTATAAGAGAAAGGTTCAATTAATTTGTTATTTGCCAAGTGGCAGCC 874

QY 723 ttgaaatccgaggggacagaaataaactgtggaatttaa 765  
Db 875 CACAAAATTAAGTAGGAGACAGAAATTTGTAATTTATCAATTTGTGA 917

RESULT 15  
SCU16731 1022 bp DNA circular VRL 07-JUN-1995  
LOCUS  
DEFINITION  
Subterranean clover stunt virus possible replication associated  
protein (SCSV2) gene, complete cds.  
ACCESSION  
U16731  
VERSION  
U16731.1 GI:571487  
KEYWORDS

SOURCE Subterranean clover stunt virus.  
ORGANISM Subterranean clover stunt virus  
REFERENCE 1 (bases 1 to 1022)  
AUTHORS Boevink, P., Chu, P.W. and Keese, P.  
TITLE Sequence of subterranean clover stunt virus DNA: affinities with the geminiviruses  
JOURNAL Virology 207 (2), 354-361 (1995)  
MEDLINE 95193233  
REFERENCE 2 (bases 1 to 1022)  
AUTHORS Boevink, P.C.  
TITLE Direct Submission  
JOURNAL Submitted (28-OCT-1994) Petra C. Boevink, C.S.I.R.O. Division of Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia  
FEATURES  
Location/Qualifiers  
1..1022  
/organism="Subterranean clover stunt virus"  
/specific\_host="Trifolium subterraneum L."  
/db\_xref="taxon:36772"  
/chromosome="component 2"  
/lab\_host="Trifolium subterraneum cv. Mt. Barker"  
82..924  
/gene="SCSV2"  
82..924  
/gene="SCSV2"  
/note="possible replication associated protein"  
/codon\_start=1  
/protein\_id="AAA68018.1"  
/db\_xref="GI:571488"  
/translation="MARRYCTFNATYETETRETSLFSDQLNLFVVGDETATTTGOK  
HLQGFVSFKNRIRLGLKKFKGNRAHWEIARGSDSONRDYCKETLISGIPVMKGS  
NKRMTETIEDPEEMQLKDPFALRCRAKKLKEEYCSYDPOKLRPWOIEHAAALMA  
EPDERSLIWYGGEGKTSFAKELIRYFWFTAGKTDQVLYMAYQAPERNIADV  
PRCSSEMMNYQAMEMKNNRVFASTKYRPVLDLIRKKVHLVIFANVAPDPTKLSDR  
IIVINC"  
stem\_loop  
join(1011..1022,1..19)  
/note="region of sequence and predicted structural  
similarity with the conserved stem-loop sequence in the  
intergenic region of the geminiviruses, which is the  
origin of replication" 285 t  
BASE COUNT 307 a 185 c 245 g  
ORIGIN

Query Match 14.2%; Score 111.2; DB 59; Length 1022;  
Best Local Similarity 49.6%; Pred. No. 5.4e-22;  
Matches 380; Conservative 0; Mismatches 368; Indels 18; Gaps 3;

QY 1 gttgggagcaggtgtgacacttcaactggccacacacacactcccaaggattccacttg 60  
Db 178 GTTGTGCGGACGAAACTGCACACTACTGGACAGAAACACCTCCAGGATTTGTATCGTTC 237

QY 61 aagaccggtcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 120  
Db 238 AAGAAAAAATTCGTCTTGTGGATTGAGAGAAATTTGGTAATC---GAGCTCACTGG 294

QY 121 gagcgcaccgtgttccgacacagacagacagcagcagcagcagcagcagcagcagcagc 180  
Db 295 GAAATTCGAGAGGAGCAGCGATTCTCAGATTCGGATTATTGTGTAAGAAAAACCCCTAAT 354

QY 181 ctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240  
Db 355 TCTGAAATTTGGGATTCGGGTCATGAGGGTTCCGAACACAGCGGAGACGATGGAGATTTAT 414

QY 241 gctgaggaacctgatgaactccgctggaagaccgccagcagcagcagcagcagcagcagc 300  
Db 415 GAAGAGATCCCGAAGAAATGCAATTTGAAGGATCCAGATCTGCTTCGATGTAAGCGG 474

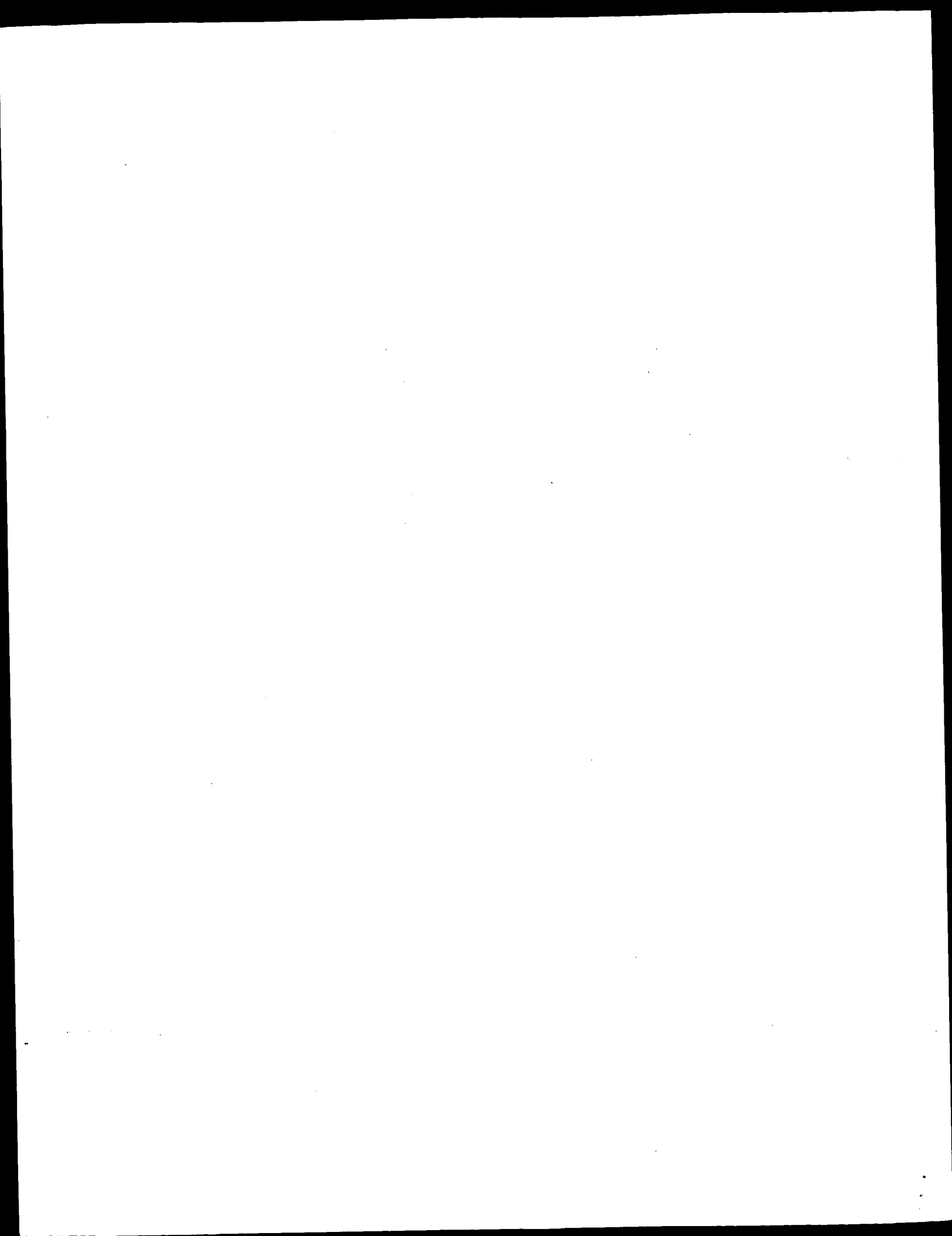
QY 301 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360  
Db 475 AAGAAATTTGAAGAGAGAAATTTGTTCTGTTATGATTTTCAGAAACTCCGTCAC----- 528

QY 361 cacaattggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420



[illegible]

Search completed: October 24, 2001, 11:42:44  
Job time: 6349 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:58 ; Search time 5701.85 Seconds  
(without alignments)  
24.415 Million cell updates/sec

Title: US-09-462-955-5  
Perfect score: 9  
Sequence: 1 ctatgatta 9

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_ov:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_ro:\*  
50: em\_sy:\*  
51: em\_sy:\*  
52: em\_un:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result | No. | Score | Query Match % | Length | DB ID       | Description         |
|--------|-----|-------|---------------|--------|-------------|---------------------|
|        |     |       |               |        |             |                     |
| C      | 1   | 9     | 100.0         | 12     | 9 AR063463  | AR063463 Sequence   |
|        | 2   | 9     | 100.0         | 22     | 10 AX069023 | AX069023 Sequence   |
|        | 3   | 9     | 100.0         | 30     | 9 AR097505  | AR097505 Sequence   |
|        | 4   | 9     | 100.0         | 32     | 9 A93671    | A93671 Sequence 7   |
| C      | 5   | 9     | 100.0         | 35     | 9 A62983    | A62983 Sequence 2   |
|        | 6   | 9     | 100.0         | 36     | 10 E22884   | E22884 DNA sequence |
|        | 7   | 9     | 100.0         | 38     | 9 AR093937  | AR093937 Sequence   |
|        | 8   | 9     | 100.0         | 38     | 10 AX081631 | AX081631 Sequence   |

Thu Oct 25 13:08:56 2001

us-09-462-955-5.ige

```

c 9 100.0 38 10 E22871 DNA sequence
10 9 100.0 39 10 AX081623
11 9 100.0 41 9 AR104937 Sequence
12 9 100.0 45 9 A39423
13 9 100.0 62 9 A39327
14 9 100.0 74 10 AX069021
15 9 100.0 81 58 AF227763
16 9 100.0 86 97 HSZ99564
17 9 100.0 94 2 CC0133579
18 9 100.0 94 2 CJE133576
19 9 100.0 101 10 I12330
20 9 100.0 101 10 I70233
21 9 100.0 103 88 AF055272
22 9 100.0 106 54 G32785
23 9 100.0 111 2 CJE133577
24 9 100.0 111 59 HIV1M00951
25 9 100.0 114 97 HUMPERNAC
26 9 100.0 117 54 G00157
27 9 100.0 120 14 S76763
28 9 100.0 121 58 AF203205
29 9 100.0 121 58 AF203214
30 9 100.0 121 93 HSPA16F10
31 9 100.0 123 3 RBHKAN2
32 9 100.0 126 3 SSISSAMB
33 9 100.0 126 93 HSAX1532
34 9 100.0 127 3 PBOKAN2
35 9 100.0 127 3 INS2MUTA
36 9 100.0 129 58 AF203217
37 9 100.0 130 3 INS43AAA
38 9 100.0 133 58 AF203212
39 9 100.0 136 54 G44046
40 9 100.0 138 59 HIVU76695
41 9 100.0 139 54 G34972
42 9 100.0 141 5 AF068581
43 9 100.0 141 5 HSPE05H04
44 9 100.0 142 3 LACSS38AA
45 9 100.0 143 1 AB005263

```

ALIGNMENTS

```

RESULT 1
LOCUS AR063463 12 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 15 from patent US 5846705.
ACCESSION AR063463
VERSION AR063463.1 GI:5992771
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Wu, R., You, L. and Soong, T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus
JOURNAL Patent: US 5846705-A 15 08-DEC-1998;
FEATURES
source Location/Qualifiers
BASE COUNT 3 a 4 c 1 g 4 t
ORIGIN

```

```

Query Match 100.0%; Score 9; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
Db 2 CTAGTATTATA 10

```

```

RESULT 2
LOCUS AX069023/c 22 bp DNA PAT 25-JAN-2001
DEFINITION Sequence 3 from Patent WO0102555.
ACCESSION AX069023
VERSION AX069023.1 GI:12578838
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Gicquel, B., Guillot, C. and Camacho, L.
TITLE Method of making and identifying attenuated microorganisms, compositions utilizing the sequences responsible for attenuation, and preparations containing attenuated microorganisms
JOURNAL Patent: WO 0102555-A 3 11-JAN-2001;
FEATURES
source Location/Qualifiers
BASE COUNT 7 a 5 c 1 g 9 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.9e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
Db 11 CTAGTATTATA 3

RESULT 3
LOCUS AR097505 30 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 33 from patent US 6071732.
ACCESSION AR097505
VERSION AR097505.1 GI:12806235
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Moore, K.L.
TITLE Tyrosylprotein sulfotransferases, nucleic acids encoding tyrosylprotein sulfotransferases, and methods of use thereof
JOURNAL Patent: US 6071732-A 33 06-JUN-2000;
FEATURES
source Location/Qualifiers
BASE COUNT 5 a 9 c 6 g 10 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 30;
Best Local Similarity 100.0%; Pred. No. 9.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
Db 4 CTAGTATTATA 12

RESULT 4
LOCUS A93671 32 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9734144.
ACCESSION A93671

```

```

VERSION      A93671.1  GI:6741859
KEYWORDS
SOURCE
ORGANISM      synthetic construct.
              synthetic construct.
              artificial sequence.
REFERENCE     1 (bases 1 to 32)
AUTHORS      Langer, G. and Yoshida, L.
TITLE        METHOD OF DETECTING THE EFFECT OF TEST SUBSTANCES USING HEN
              UROKINASE
JOURNAL
PATENT       WO 9734144-A 7 18-SEP-1997;
              SCHERING AG (DE); LANGER GERNOT (DE)
FEATURES
source       1..32
              /organism="synthetic construct"
              /db_xref="taxon:32630"
BASE COUNT   5 a      7 c      11 g      9 t
ORIGIN
Query Match  100.0%; Score 9; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 9.1e+04;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
    |||||
Db 12 CTAGTATTA 20

RESULT 5
LOCUS      A62983
DEFINITION Sequence 2 from Patent WO9720056.
ACCESSION  A62983
VERSION     A62983.1  GI:3716855
KEYWORDS
SOURCE     pea.
ORGANISM   Pisum sativum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Pisum.
REFERENCE
AUTHORS    Gray, J.C., Sandhu, J.S. and Webster, C.I.
TITLE      ENHANCER-INCREASED GENE EXPRESSION IN PLANTS
JOURNAL    Patent: WO 9720056-A 2 05-JUN-1997;
            CAMBRIDGE ADVANCED TECH (GB)
COMMENT    Other publication AU 7635696 19970619.
FEATURES
source     1..35
            /organism="Pisum sativum"
            /db_xref="taxon:3888"
            /clone_lib="GENOMIC, LAMBDA EMBL3"
            /clone_lib="(X)SEQUENCE DESCRIPTION:SEQ ID NO: 2:"
BASE COUNT 17 a      4 c      2 g      12 t
ORIGIN
Query Match  100.0%; Score 9; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 8.9e+04;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
    |||||
Db 11 CTAGTATTA 19

RESULT 6
LOCUS      E22884
DEFINITION DNA sequence increasing promoter activity.
ACCESSION  E22884
VERSION     E22884.1  GI:13024162
KEYWORDS    JP 1999075843-A/21.
SOURCE     unidentified.

us-09-462-955-5.rge
VERSION      A93671.1  GI:6741859
KEYWORDS
SOURCE
ORGANISM      unclassified.
              unclassified.
              1 (bases 1 to 36)
              Toshihiro, Y.
              DNA sequence increasing promoter activity
              Patent: JP 1999075843-A 21 23-MAR-1999;
              KIRIN BREWERY CO LTD
COMMENT       OS Unidentified
              PN JP 1999075843-A/21
              PD 23-MAR-1999
              PF 29-AUG-1997 JP 1997234995
              PR
              PI TOSHIHIRO YONEDA
              PC C12N15/09,C12N1/19,C12P21/02//(C12N15/09,C12R1:72),(C12N1/19,
              PC C12R1:72),
              PC (C12P21/02,C12R1:72),C12N15/00,(C12N15/00,C12R1:72) CC
              Strandedness: Single;
              CC Topology: Linear;
              FH Key Location/Qualifiers
              FT source 1..36
              FT /organism="Unidentified"
              FT Location/Qualifiers
              source 1..36
              /organism="unidentified"
              /db_xref="taxon:32644"
BASE COUNT   13 a      8 c      3 g      12 t
ORIGIN
Query Match  100.0%; Score 9; DB 10; Length 36;
Best Local Similarity 100.0%; Pred. No. 8.9e+04;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
    |||||
Db 26 CTAGTATTA 34

RESULT 7
LOCUS      AR093937/c
DEFINITION Sequence 41 from patent US 6001590.
ACCESSION  AR093937
VERSION     AR093937.1  GI:10020682
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE
AUTHORS    Komeda, T., Suda, H., Tamai, Y., Iwamatsu, A., Kato, N. and Sakai, Y.
TITLE      Promoter and terminator sequences of formate dehydrogenase gene of
              Candida boidinii
JOURNAL    Patent: US 6001590-A 41 14-DEC-1999;
FEATURES
source     1..38
            Location/Qualifiers
BASE COUNT 12 a      8 c      5 g      13 t
ORIGIN
Query Match  100.0%; Score 9; DB 9; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
    |||||
Db 18 CTAGTATTA 10

RESULT 8
LOCUS      AX081631
DEFINITION Sequence 136 from Patent WO0109350.

```

Thu Oct 25 13:08:56 2001

us-09-462-955-5.rge

```

Db 18 CTAGTATTA 10

RESULT 10
AX081623 39 bp DNA PAT 27-FEB-2001
LOCUS
AX081623
DEFINITION
Sequence 128 from Patent WO0109350.
ACCESSION
AX081623
VERSION
AX081623.1 GI:13170448
KEYWORDS
synthetic construct.
ORGANISM
synthetic construct
artificial sequence.
REFERENCE
1 (bases 1 to 39)
AUTHORS
Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
TITLE
Genetically engineered bleb vaccine
JOURNAL
Patent: WO 0109350-A 128 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
Location/Qualifiers
source
1..39
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="BAD 03-2 primer"
BASE COUNT 6 a 14 c 10 g 9 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 39;
Best Local Similarity 100.0%; Pred. No. 8.7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
|||||
Db 18 CTAGTATTA 27

RESULT 11
AX081623 41 bp DNA PAT 14-FEB-2001
LOCUS
AX081623
DEFINITION
Sequence 106 from patent US 6096321.
ACCESSION
AX081623
VERSION
AX081623.1 GI:12818534
KEYWORDS
Unknown.
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 41)
AUTHORS
Girardeau,J., Martin,C., Mechin,M., Der Vartanian,M. and
Bousquet,F.
TITLE
Clp3 subunit of CS31A protein capsule containing heterologous
peptides
JOURNAL
Patent: US 6096321-A 106 01-AUG-2000;
FEATURES
Location/Qualifiers
source
1..41
/organism="unknown"
BASE COUNT 11 a 3 c 10 g 17 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 8.6e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
|||||
Db 1 CTAGTATTA 9

RESULT 12
AX39423 45 bp DNA PAT 05-MAR-1997
LOCUS
AX39423
DEFINITION
Sequence 42 from Patent WO9414967.

```

```

ACCESSION AX081631
VERSION AX081631.1 GI:13170456
KEYWORDS
synthetic construct.
ORGANISM
synthetic construct
artificial sequence.
REFERENCE
1 (bases 1 to 38)
AUTHORS
Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
TITLE
Genetically engineered bleb vaccine
JOURNAL
Patent: WO 0109350-A 136 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
Location/Qualifiers
source
1..38
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="BAD20 primer"
BASE COUNT 7 a 14 c 6 g 11 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
|||||
Db 18 CTAGTATTA 26

RESULT 9
E22871/c 38 bp DNA PAT 07-FEB-2001
LOCUS
E22871
DEFINITION
DNA sequence increasing promoter activity.
ACCESSION
E22871
VERSION
E22871.1 GI:13024149
KEYWORDS
JP 1999075843-A/8.
SOURCE
unidentified.
ORGANISM
unclassified.
REFERENCE
Toshihiro,Y.
AUTHORS
DNA sequence increasing promoter activity
TITLE
Patent: JP 1999075843-A 8 23-MAR-1999;
JOURNAL
KIRIN BREWERY CO LTD
COMMENT
OS Unidentified
PN JP 1999075843-A/8
PD 23-MAR-1999
PF 29-AUG-1997 JP 1997234995
PR
PI TOSHIHIRO YONEDA
PC C12N15/09,C12N1/19,C12P21/02,((C12N15/09,C12R1:72),(C12N1/19,
PC C12R1:72),
PC C12P21/02,C12R1:72),C12N15/00,(C12N15/00,C12R1:72) CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..38
/organism="unidentified"
FEATURES
Location/Qualifiers
source
1..38
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 12 a 8 c 5 g 13 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
|||||

```

ACCESSION A39423  
 VERSION A39423.1 GI:2295756  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 45)  
 AUTHORS Girardeau, J., Martin, C., Mechin, M., Der, V. M. and Bousquet, F.  
 TITLE CS31A PROTEIN CAPSULE SUBUNIT MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A PROTEIN CAPSULE COMPRISING SAID SUBUNIT, MICRO-ORGANISMS HAVING SUCH SUBUNITS ON THEIR OUTER MEMBRANE, AND METHODS FOR PREPARING AND USING SAME  
 JOURNAL Patent: WO 9414967-A 42 07-JUL-1994;  
 COMMENT AGRONOMIQUE INST NAT RECH (FR)  
 Other publication AU 6808794 940719  
 Other publication FR 2699538 940624.  
 FEATURES  
 source 1..45  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 12 a 4 c 11 g 18 t  
 ORIGIN  
 Query Match 100.0%; Score 9; DB 9; Length 45;  
 Best Local Similarity 100.0%; Pred. No. 8.5e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ctagtattta 9  
 |||||  
 Db 1 CTAGTATTTA 9  
 RESULT 13  
 A39327  
 LOCUS A39327 62 bp DNA PAT 05-MAR-1997  
 DEFINITION Sequence 9 from Patent WO9414969.  
 ACCESSION A39327  
 VERSION A39327.1 GI:2295680  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 62)  
 AUTHORS Adams, S.E., Burns, R.N. and Richardson, S.M.  
 TITLE NOVEL PROTEINACEOUS PARTICLES  
 JOURNAL Patent: WO 9414969-A 9 07-JUL-1994;  
 BRITISH BIO TECHNOLOGY (GB)  
 COMMENT Other publication AU 5711294 940719  
 Other publication JP 8505788T 960625.  
 FEATURES  
 source 1..62  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 23 a 12 c 10 g 17 t  
 ORIGIN  
 Query Match 100.0%; Score 9; DB 9; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 7.9e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ctagtattta 9  
 |||||  
 Db 1 CTAGTATTTA 9  
 RESULT 14  
 AX069021  
 LOCUS AX069021 74 bp DNA PAT 25-JAN-2001  
 DEFINITION Sequence 1 from Patent WO0102555.  
 ACCESSION AX069021  
 VERSION AX069021.1 GI:12578836

KEYWORDS  
 SOURCE Mycobacterium tuberculosis.  
 ORGANISM Mycobacterium tuberculosis  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
 Mycobacterium; Mycobacterium tuberculosis complex.  
 REFERENCE 1 (bases 1 to 74)  
 AUTHORS Gicquel, B., Guilhot, C. and Camacho, L.  
 TITLE Method of making and identifying attenuated microorganisms, compositions utilizing the sequences responsible for attenuation, and preparations containing attenuated microorganisms  
 JOURNAL Patent: WO 0102555-A 1 11-JAN-2001;  
 INSTITUT PASTEUR (FR)  
 FEATURES  
 source 1..74  
 /organism="Mycobacterium tuberculosis"  
 /db\_xref="taxon:1773"  
 BASE COUNT 21 a 9 c 8 g 16 t 20 others  
 ORIGIN  
 Query Match 100.0%; Score 9; DB 10; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 7.6e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ctagtattta 9  
 |||||  
 Db 64 CTAGTATTTA 72  
 RESULT 15  
 AF227763  
 LOCUS AF227763 81 bp RNA VRL 12-JAN-2001  
 DEFINITION Hepatitis C virus isolate C10 envelope protein 2 (E2) gene, partial cds.  
 ACCESSION AF227763  
 VERSION AF227763.1 GI:8099446  
 KEYWORDS  
 SOURCE Hepatitis C virus.  
 ORGANISM Hepatitis C virus.  
 REFERENCE 1 (bases 1 to 81)  
 AUTHORS Ross, R.S., Viazov, S., Gross, T., Hofmann, F., Seipp, H.M. and Roggendorf, M.  
 TITLE Transmission of the hepatitis C, virus from a patient to an anesthesiology assistant to five patients  
 JOURNAL N. Engl. J. Med. 343 (25), 1851-1854 (2000)  
 MEDLINE 21004055  
 PUBMED 11117977  
 REFERENCE 2 (bases 1 to 81)  
 AUTHORS Ross, S., Viazov, S. and Roggendorf, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JAN-2000) Institute of Virology, Essen University Hospital, Hufelandstr. 55, Essen D-45122, Germany  
 FEATURES  
 source 1..81  
 /organism="Hepatitis C virus"  
 /isolate="C10"  
 /db\_xref="taxon:11103"  
 /note="genotype: 1a"  
 <1..>81  
 /gene="E2"  
 <1..>81  
 /gene="E2"  
 /note="hypervariable region I; HVR1"  
 /codon\_start=1  
 /product="envelope protein 2"  
 /protein\_id="AA072123.1"  
 /db\_xref="GI:8099447"  
 /translation="HPTVTGGSAAYTSSITRLTPGAKQN"  
 BASE COUNT 21 a 29 c 18 g 13 t  
 ORIGIN

Thu Oct 25 13:08:56 2001

Query Match 100.0%; Score 9; DB 58; Length 81;  
Best Local Similarity 100.0%; Pred. No. 7.5e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9  
|||||  
Db 41 CTAGTATTTA 49

Search completed: October 24, 2001, 11:42:58  
Job time: 6363 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:55 ; Search time 5701.85 Seconds  
(without alignments)  
24.415 Million cell updates/sec

Title: US-09-462-955-4  
Perfect score: 9  
Sequence: 1 taatactag 9

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_bal:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_om:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pl1:\*

13: gb\_pl2:\*

14: gb\_pl3:\*

15: gb\_pl4:\*

16: em\_bal:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_htgo\_hum:\*

20: em\_htgo\_inv:\*

21: em\_htgo\_rsd:\*

22: em\_htg\_hum1:\*

23: em\_htg\_hum2:\*

24: em\_htg\_hum3:\*

25: em\_htg\_hum4:\*

26: em\_htg\_hum5:\*

27: em\_htg\_hum6:\*

28: em\_htg\_hum7:\*

29: em\_htg\_hum8:\*

30: em\_htg\_inv1:\*

31: em\_htg\_inv2:\*

32: em\_htg\_other:\*

33: em\_htg\_rsd:\*

34: em\_hum1:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_om:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_rsd:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_v1:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_v1:\*

59: gb\_v12:\*

60: gb\_htg1:\*

61: gb\_htg2:\*

62: gb\_htg3:\*

63: gb\_htg4:\*

64: gb\_htg5:\*

65: gb\_htg6:\*

66: gb\_htg7:\*

67: gb\_htg8:\*

68: gb\_htg9:\*

69: gb\_htg10:\*

70: gb\_htg11:\*

71: gb\_htg12:\*

72: gb\_htg13:\*

73: gb\_htg14:\*

74: gb\_htg15:\*

75: gb\_htg16:\*

76: gb\_htg17:\*

77: gb\_htg18:\*

78: gb\_htg19:\*

79: gb\_htg20:\*

80: gb\_htg21:\*

81: gb\_htg22:\*

82: gb\_htg23:\*

83: gb\_htg24:\*

84: gb\_htg25:\*

85: gb\_pr1:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_rsd:\*

95: gb\_rsd2:\*

96: gb\_in4:\*

97: gb\_pr10:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | %     | Query | Match | Length   | DB | ID | Description         |
|--------|-----|-------|-------|-------|-------|----------|----|----|---------------------|
| C      | 1   | 9     | 100.0 | 12    | 9     | AR063463 |    |    | AR063463 Sequence   |
| C      | 2   | 9     | 100.0 | 22    | 10    | AX069023 |    |    | AX069023 Sequence   |
| C      | 3   | 9     | 100.0 | 30    | 9     | AR097505 |    |    | AR097505 Sequence   |
| C      | 4   | 9     | 100.0 | 32    | 9     | A93671   |    |    | A93671 Sequence 7   |
| C      | 5   | 9     | 100.0 | 35    | 9     | A62983   |    |    | A62983 Sequence 2   |
| C      | 6   | 9     | 100.0 | 36    | 10    | E22884   |    |    | E22884 DNA sequence |
| C      | 7   | 9     | 100.0 | 38    | 9     | AR093937 |    |    | AR093937 Sequence   |
| C      | 8   | 9     | 100.0 | 38    | 10    | AX081631 |    |    | AX081631 Sequence   |

9 100.0 38 10 E22871 DNA sequenc  
c 10 9 100.0 39 10 AX081623 Sequence  
c 11 9 100.0 41 9 AR104937 Sequence  
c 12 9 100.0 45 9 A39423 Sequence 42  
c 13 9 100.0 62 9 A39327 Sequence 9  
c 14 9 100.0 74 10 AX069021 Sequence  
c 15 9 100.0 81 58 AF227763 Hepatitis  
c 16 9 100.0 86 97 HS299564  
c 17 9 100.0 94 2 CC0133579  
c 18 9 100.0 94 2 CJE133576  
c 19 9 100.0 101 10 I12330  
c 20 9 100.0 101 10 I70233 Sequence 14  
c 21 9 100.0 103 88 AF055272 Homo sapi  
c 22 9 100.0 106 54 G32785  
c 23 9 100.0 111 2 CJE133577  
c 24 9 100.0 111 59 HIV1M009S1  
c 25 9 100.0 114 97 HUMPSRNAC  
c 26 9 100.0 117 54 G00157  
c 27 9 100.0 120 14 S76763  
c 28 9 100.0 121 58 AF203205  
c 29 9 100.0 121 58 AF203214  
c 30 9 100.0 121 93 HSPAL6F10  
c 31 9 100.0 123 3 RBHKAN2  
c 32 9 100.0 126 3 SSISSAMB  
c 33 9 100.0 126 93 HSAOX1S32  
c 34 9 100.0 127 3 PR0KAN2  
c 35 9 100.0 127 58 AF203217  
c 36 9 100.0 129 3 INS2MUTA  
c 37 9 100.0 130 3 INS43AAA  
c 38 9 100.0 133 58 AF203212  
c 39 9 100.0 136 54 G44046  
c 40 9 100.0 138 59 HIVU76695  
c 41 9 100.0 139 54 G34372  
c 42 9 100.0 141 5 AF068581  
c 43 9 100.0 141 54 HSP05H04  
c 44 9 100.0 142 3 LACS38AA  
c 45 9 100.0 143 1 AB005263

ALIGNMENTS

RESULT 1  
AR063463/c  
LOCUS AR063463 12 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 15 from patent US 5846705.  
ACCESSION AR063463  
VERSION AR063463.1 GI:5992771  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 12)  
AUTHORS Wu, R., You, L. and Soong, T.  
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus  
JOURNAL Patent: US 5846705-A 15 08-DEC-1998;  
FEATURES Location/Qualifiers  
source 1..12  
BASE COUNT 3 a 4 c 1 g 4 t  
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 taactag 9  
Db 10 TAATACTAG 2

RESULT 2  
AX069023  
LOCUS AX069023 22 bp DNA PAT 25-JAN-2001  
DEFINITION Sequence 3 from Patent WO0102555.  
ACCESSION AX069023  
VERSION AX069023.1 GI:12578838  
KEYWORDS  
SOURCE Mycobacterium tuberculosis.  
ORGANISM Mycobacterium tuberculosis.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium; Mycobacterium tuberculosis complex.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Gicquel, B., Guilhot, C. and Camacho, L.  
TITLE Method of making and identifying attenuated microorganisms, compositions utilizing the sequences responsible for attenuation, and preparations containing attenuated microorganisms  
JOURNAL Patent: WO 0102555-A 3 11-JAN-2001;  
INSTITUT PASTEUR (FR)  
FEATURES Location/Qualifiers  
source 1..22  
/organism="Mycobacterium tuberculosis"  
/db\_xref="taxon:1773" 9 t  
BASE COUNT 7 a 5 c 1 g  
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 22;  
Best Local Similarity 100.0%; Pred. No. 9.9e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 taactag 9  
Db 3 TAATACTAG 11  
RESULT 3  
AR097505/c  
LOCUS AR097505 30 bp DNA PAT 14-FEB-2001  
DEFINITION Sequence 33 from patent US 6071732.  
ACCESSION AR097505  
VERSION AR097505.1 GI:12806235  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Moore, K.L.  
TITLE Tyrosylprotein sulfotransferases, nucleic acids encoding tyrosylprotein sulfotransferases, and methods of use thereof  
JOURNAL Patent: US 6071732-A 33 06-JUN-2000;  
FEATURES Location/Qualifiers  
source 1..30  
/organism="unknown" 10 t  
BASE COUNT 5 a 9 c 6 g  
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 30;  
Best Local Similarity 100.0%; Pred. No. 9.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 taactag 9  
Db 12 TAATACTAG 4  
RESULT 4  
A93671/c  
LOCUS A93671 32 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 7 from Patent WO9734144.  
ACCESSION A93671



Thu Oct 25 13:08:49 2001

us-09-462-955-4.rge

```

ACCESSION AX081631
VERSION AX081631.1 GI:13170456
KEYWORDS
SOURCE
ORGANISM synthetic construct.
          synthetic construct
          artificial sequence.
REFERENCE 1 (bases 1 to 38)
AUTHORS Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
          Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
TITLE Genetically engineered bleb vaccine
JOURNAL Patent: WO 0109350-A 136 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
  source
    Location/Qualifiers
      1..38
      /organism="synthetic construct"
      /db_xref="taxon:32630"
      /note="BAD20 primer"
BASE COUNT 7 a 14 c 6 g 11 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taactag 9
|||||
Db 26 TAATACTAG 18

RESULT 9
E22871 38 bp DNA PAT 07-FEB-2001
LOCUS
DEFINITION DNA sequence increasing promoter activity.
ACCESSION E22871
VERSION E22871.1 GI:13024149
KEYWORDS JP 1999075843-A/8.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 38)
AUTHORS Toshihiro,Y.
TITLE DNA sequence increasing promoter activity
JOURNAL Patent: JP 1999075843-A 8 23-MAR-1999;
          KIRIN BREWERY CO LTD
COMMENT OS Unidentified
          PN JP 1999075843-A/8
          PD 23-MAR-1999
          PF 29-AUG-1997 JP 1997234995
          PR
PI TOSHIHIRO YONEDA
PC C12N15/09,C12N1/19,C12P21/02,/(C12N15/09,C12R1:72),(C12N1/19,
PC C12R1:72),
PC C12R1:72),
PC (C12P21/02,C12R1:72),C12N15/00,(C12N15/00,C12R1:72) CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..38
   /organism="Unidentified".
FEATURES
  source
    Location/Qualifiers
      1..38
      /organism="unidentified"
      /db_xref="taxon:32644"
BASE COUNT 12 a 8 c 5 g 13 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taactag 9
|||||

```

```

Db 10 TAATACTAG 18

RESULT 10
AX081623/c 39 bp DNA PAT 27-FEB-2001
LOCUS
DEFINITION Sequence 128 from Patent WO0109350.
ACCESSION AX081623
VERSION AX081623.1 GI:13170448
KEYWORDS synthetic construct.
SOURCE synthetic construct
          artificial sequence.
ORGANISM
REFERENCE 1 (bases 1 to 39)
AUTHORS Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
          Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
TITLE Genetically engineered bleb vaccine
JOURNAL Patent: WO 0109350-A 128 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
  source
    Location/Qualifiers
      1..39
      /organism="synthetic construct"
      /db_xref="taxon:32630"
      /note="BAD 03-2 primer"
BASE COUNT 6 a 14 c 10 g 9 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 39;
Best Local Similarity 100.0%; Pred. No. 8.7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taactag 9
|||||
Db 27 TAATACTAG 19

RESULT 11
AR104937 41 bp DNA PAT 14-FEB-2001
LOCUS
DEFINITION Sequence 106 from patent US 6096321.
ACCESSION AR104937
VERSION AR104937.1 GI:12818534
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Girardeau,J., Martin,C., Mechin,M., Der Vartanian,M. and
          Bousquet,F.
TITLE C1pG subunit of CS31A protein capsule containing heterologous
          peptides
JOURNAL Patent: US 6096321-A 106 01-AUG-2000;
FEATURES
  source
    Location/Qualifiers
      1..41
      /organism="unknown"
BASE COUNT 11 a 3 c 10 g 17 t
ORIGIN
Query Match 100.0%; Score 9; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 8.6e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taactag 9
|||||
Db 9 TAATACTAG 1

RESULT 12
A39423/c 45 bp DNA PAT 05-MAR-1997
LOCUS
DEFINITION Sequence 42 from Patent WO9414967.

```

ACCESSION A39423  
 VERSION A39423.1 GI:2295756  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unclassified.  
 unclassified.  
 unclassified.  
 1 (bases 1 to 45)  
 Girardeau, J., Martin, C., Mechin, M., Der, V.M. and Bousquet, F.  
 CS31A PROTEIN CAPSULE SUBUNIT MODIFIED BY AT LEAST ONE HETEROLOGOUS  
 PEPTIDE, CS31A PROTEIN CAPSULE COMPRISING SAID SUBUNIT,  
 MICRO-ORGANISMS HAVING SUCH SUBUNITS ON THEIR OUTER MEMBRANE, AND  
 METHODS FOR PREPARING AND USING SAME  
 Patent: WO 9414967-A 42 07-JUL-1994;  
 AGRONOMIQUE INST NAT RECH (FR)  
 Other publication AU 6808794 940719  
 Other publication FR 2699538 940624.  
 FEATURES  
 source  
 1..45  
 Location/Qualifiers  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 12 a 4 c 11 g 18 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 45;  
 Best Local Similarity 100.0%; Pred. No. 8.5e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 |||||  
 Db 9 TAATACTAG 1

RESULT 13  
 A39327/c  
 LOCUS A39327 62 bp DNA  
 DEFINITION Sequence 9 from Patent WO9414969.  
 ACCESSION A39327  
 VERSION A39327.1 GI:2295680  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unidentified.  
 unclassified.  
 unclassified.  
 1 (bases 1 to 62)  
 Adams, S.E., Burns, R.N. and Richardson, S.M.  
 NOVEL PROTEINACEOUS PARTICLES  
 Patent: WO 9414969-A 9 07-JUL-1994;  
 BRITISH BIO TECHNOLOGY (GB)  
 Other publication AU 5711294 940719  
 Other publication JP 8505768T 960625.  
 FEATURES  
 source  
 1..62  
 Location/Qualifiers  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 23 a 12 c 10 g 17 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 7.9e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 |||||  
 Db 9 TAATACTAG 1

RESULT 14  
 AX069021/c  
 LOCUS AX069021 74 bp DNA  
 DEFINITION Sequence 1 from Patent WO0102555.  
 ACCESSION AX069021  
 VERSION AX069021.1 GI:12578836

KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium tuberculosis.  
 Mycobacterium tuberculosis  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
 Mycobacterium; Mycobacterium tuberculosis complex.  
 1 (bases 1 to 74)  
 Gicquel, B., Guilhot, C. and Camacho, L.  
 Method of making and identifying attenuated microorganisms,  
 compositions utilizing the sequences responsible for attenuation,  
 and preparations containing attenuated microorganisms  
 Patent: WO 0102555-A 1 11-JAN-2001;  
 INSTITUT PASTEUR (FR)  
 FEATURES  
 source  
 1..74  
 Location/Qualifiers  
 /organism="Mycobacterium tuberculosis"  
 /db\_xref="taxon:1773"  
 BASE COUNT 21 a 9 c 8 g 16 t 20 others  
 ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 7.6e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 |||||  
 Db 72 TAATACTAG 64

RESULT 15  
 AF227763/c  
 LOCUS AF227763 81 bp RNA  
 DEFINITION Hepatitis C virus isolate C10 envelope protein 2 (E2) gene, partial cds.  
 ACCESSION AF227763  
 VERSION AF227763.1 GI:8099446  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Hepatitis C virus.  
 Hepatitis C virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Hepacivirus.  
 1 (bases 1 to 81)  
 Ross, R.S., Viazov, S., Gross, T., Hofmann, F., Seipp, H.M. and  
 Roggendorf, M.  
 Transmission of the hepatitis C virus from a patient to an  
 anesthesiology assistant to five patients  
 N. Engl. J. Med. 343 (25), 1851-1854 (2000)  
 21004055  
 MEDLINE  
 PUBMED 11117977  
 REFERENCE  
 2 (bases 1 to 81)  
 Ross, R.S., Viazov, S. and Roggendorf, M.  
 Direct Submission  
 Submitted (21-JAN-2000) Institute of Virology, Essen University  
 Hospital, Hufelandstr. 55, Essen D-45122, Germany  
 FEATURES  
 source  
 1..81  
 Location/Qualifiers  
 /organism="Hepatitis C virus"  
 /isolate="C10"  
 /db\_xref="taxon:11103"  
 /note="genotype: 1a"  
 <1..>81  
 /gene="E2"  
 <1..>81  
 /gene="E2"  
 /note="hypervariable region 1; HVRI"  
 /codon\_start=1  
 /product="envelope protein 2"  
 /protein\_id="AAF72123.1"  
 /db\_xref="GI:8099447"  
 /translation="HTHTGGSAAYTSSITRLTTPGAKQN"  
 BASE COUNT 21 a 29 c 18 g 13 t  
 ORIGIN

Thu Oct 25 13:08:49 2001

us-09-462-955-4.rge

Query Match 100.0%; Score 9; DB 58; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 7.5e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 taatactag 9  
 |||||  
 Db 49 TAATACTAG 41

Search completed: October 24, 2001, 11:42:58  
 Job time: 6363 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:53 : Search time 5701.85 Seconds  
(without alignments)  
24.415 Million cell updates/sec

Title: US-09-462-955-3

Perfect score: 9

Sequence: 1 taatattac 9

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_ba3.\*

4: gb\_in1.\*

5: gb\_in2.\*

6: gb\_in3.\*

7: gb\_om.\*

8: gb\_ov.\*

9: gb\_pat1.\*

10: gb\_pat2.\*

11: gb\_ph.\*

12: gb\_pl1.\*

13: gb\_pl2.\*

14: gb\_pl3.\*

15: gb\_pl4.\*

16: gb\_ba1.\*

17: gb\_ba2.\*

18: em\_fun.\*

19: em\_htgo\_hum.\*

20: em\_htgo\_inv.\*

21: em\_htgo\_rdd.\*

22: em\_htg\_hum1.\*

23: em\_htg\_hum2.\*

24: em\_htg\_hum3.\*

25: em\_htg\_hum4.\*

26: em\_htg\_hum5.\*

27: em\_htg\_hum6.\*

28: em\_htg\_hum7.\*

29: em\_htg\_hum8.\*

30: em\_htg\_inv1.\*

31: em\_htg\_inv2.\*

32: em\_htg\_other.\*

33: em\_hum1.\*

34: em\_hum2.\*

35: em\_hum3.\*

36: em\_hum4.\*

37: em\_hum5.\*

38: em\_hum6.\*

39: em\_hum7.\*

40: em\_in.\*

41: em\_om.\*

42: em\_or.\*

43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_v1.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_v11.\*  
59: gb\_v12.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rod.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query |        |    | ID       | Description        |
|------------|-------|-------|--------|----|----------|--------------------|
|            |       | Match | Length | DB |          |                    |
| 1          | 9     | 100.0 | 17     | 10 | 121591   | 121591 Sequence 3  |
| 2          | 9     | 100.0 | 20     | 10 | 119565   | 119565 Sequence 3  |
| 3          | 9     | 100.0 | 20     | 10 | 176521   | 176521 Sequence 1  |
| 4          | 9     | 100.0 | 20     | 10 | 176526   | 176526 Sequence 6  |
| C 5        | 9     | 100.0 | 21     | 9  | AR102784 | AR102784 Sequence  |
| C 6        | 9     | 100.0 | 21     | 10 | E23427   | E23427 Oligonucleo |
| C 7        | 9     | 100.0 | 21     | 10 | E23436   | E23436 Oligonucleo |
| C 8        | 9     | 100.0 | 21     | 10 | E23445   | E23445 Oligonucleo |

9 100.0 21 10 I76530  
 10 100.0 24 10 E26401  
 11 100.0 24 10 E26401  
 12 100.0 26 10 AX073421  
 13 100.0 27 9 AX027097  
 14 100.0 30 97 MTHSNDU7M  
 15 100.0 34 9 A49364  
 16 100.0 35 9 A62540  
 17 100.0 35 9 AR104880  
 18 100.0 37 9 AR070909  
 19 100.0 55 10 E23318  
 20 100.0 58 58 BCTSTLPA  
 21 100.0 60 9 A84543  
 22 100.0 60 58 BCTSTLPP  
 23 100.0 72 9 AR006900  
 24 100.0 72 9 AR110924  
 25 100.0 72 10 I74840  
 26 100.0 75 6 DSNTSR4  
 27 100.0 75 6 DSNTSR4  
 28 100.0 76 7 TIOSINE  
 29 100.0 76 7 TVU14602  
 30 100.0 80 97 HSU14601  
 31 100.0 80 97 HUMSINE  
 32 100.0 90 3 STABLAZA  
 33 100.0 95 5 AF166131  
 34 100.0 98 5 AF166132  
 35 100.0 98 54 G37851  
 36 100.0 99 93 HSA241951  
 37 100.0 99 93 HSA241951  
 38 100.0 101 5 AF166133  
 39 100.0 103 54 G32992  
 40 100.0 104 92 HS87H7F  
 41 100.0 105 53 CNS0740W  
 42 100.0 109 54 G19822  
 43 100.0 110 54 G37869  
 44 100.0 110 58 BCU57490  
 45 100.0 111 54 G38825

ALIGNMENTS

RESULT 1  
 I21591  
 LOCUS I21591 17 bp DNA PAT 07-OCT-1996  
 DEFINITION Sequence 3 from patent US 5521302.  
 ACCESSION I21591  
 VERSION I21591.1 GI:1601945  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 17)  
 AUTHORS Cook,P.D.  
 TITLE Process for preparing oligonucleotides having chiral phosphorus linkages  
 JOURNAL Patent: US 5521302-A 3 28-MAY-1996;  
 FEATURES Location/Qualifiers  
 source 1..17  
 /organism="unknown"  
 BASE COUNT 6 a 3 c 2 g 6 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 taattattac 9  
 Db 7 TAATATTAC 15

RESULT 2  
 I19565  
 LOCUS I19565 20 bp DNA PAT 07-OCT-1996  
 DEFINITION Sequence 3 from patent US 5506212.  
 ACCESSION I19565  
 VERSION I19565.1 GI:1599920  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Hoke,G. and Cook,P.D.  
 TITLE Oligonucleotides with substantially chirally pure phosphorothioate linkages  
 JOURNAL Patent: US 5506212-A 3 09-APR-1996;  
 FEATURES Location/Qualifiers  
 source 1..20  
 /organism="unknown"  
 BASE COUNT 6 a 3 c 3 g 8 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 taattattac 9  
 Db 10 TAATATTAC 18

RESULT 3  
 I76521  
 LOCUS I76521 20 bp DNA PAT 03-APR-1998  
 DEFINITION Sequence 1 from patent US 5691461.  
 ACCESSION I76521  
 VERSION I76521.1 GI:3012675  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Ecker,D.J. and Hoke,G.D.  
 TITLE Oligonucleotides inhibiting candida germ tube formation  
 JOURNAL Patent: US 5691461-A 1 25-NOV-1997;  
 FEATURES Location/Qualifiers  
 source 1..20  
 /organism="unknown"  
 BASE COUNT 6 a 3 c 3 g 8 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 taattattac 9  
 Db 10 TAATATTAC 18

RESULT 4  
 I76526  
 LOCUS I76526 20 bp DNA PAT 03-APR-1998  
 DEFINITION Sequence 6' from patent US 5691461.  
 ACCESSION I76526  
 VERSION I76526.1 GI:3012680  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 20)  
 AUTHORS Ecker,D.J. and Hoke,G.D.



TITLE Oligonucleotides inhibiting candida germ tube formation  
JOURNAL Patent: US 5691461-A 6 25-NOV-1997;

FEATURES  
Source Location/Qualifiers  
1..20  
/organism="unknown"  
BASE COUNT 6 a 3 c 3 g 8 t  
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
|||||||  
Db 10 TAATATTAC 18

RESULT 5  
AR102784/c  
LOCUS 21 bp DNA PAT 14-FEB-2001  
DEFINITION Sequence 11 from patent US 6087104.  
ACCESSION AR102784  
VERSION AR102784.1 GI:12814372  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 21)  
AUTHORS Yamada,S., Venkateswaran,K. and Ohashi,E.  
TITLE Oligonucleotides for detection of Bacillus cereus group bacteria harmful to mammals, and method of detection with the oligonucleotides  
JOURNAL Patent: US 6087104-A 11 11-JUL-2000;  
FEATURES Location/Qualifiers  
Source 1..21  
/organism="unknown"

BASE COUNT 9 a 3 c 3 g 6 t  
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
|||||||  
Db 13 TAATATTAC 5

RESULT 6  
E23427/c  
LOCUS 21 bp DNA PAT 07-FEB-2001  
DEFINITION Oligonucleotide for detecting Bacillus cereus and method of detection therewith.  
ACCESSION E23427  
VERSION E23427.1 GI:13024403  
KEYWORDS JP 1999004691-A/8.  
SOURCE unidentified.  
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 21)  
AUTHORS Shiochi,Y.K.V.V. and Ohashi.  
TITLE Oligonucleotide for detecting Bacillus cereus and method of detection therewith  
JOURNAL Patent: JP 1999004691-A 8 12-JAN-1999;  
COMMENT NIPPON SUISAN KAISHA LTD  
OS Unidentified  
PN JP 1999004691-A/8  
PD 12-JAN-1999  
PF 29-SEP-1997 JP 1997264057  
PR

PI SHOICHI YAMADA, KASUTORI VENKATESWARAN,EIJI OHASHI PC

C12N15/09,C12Q1/04,C12Q1/68//((C12N15/09,C12R1:085),(C12Q1/04,PC C12R1:085)),  
PC C12N15/00,(C12N15/00,C12R1:085)  
CC Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..21  
FT /organism='Unidentified'.

FEATURES  
source Location/Qualifiers  
1..21  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 9 a 3 c 3 g 6 t  
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
|||||||  
Db 13 TAATATTAC 5

RESULT 7  
E23436/c  
LOCUS 21 bp DNA PAT 07-FEB-2001  
DEFINITION Oligonucleotide for detecting Bacillus thuringiensis and method of detection therewith.  
ACCESSION E23436  
VERSION E23436.1 GI:13024412  
KEYWORDS JP 1999004692-A/8.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Shiochi,Y.K.V.V. and Ohashi.  
TITLE Oligonucleotide for detecting Bacillus thuringiensis and method of detection therewith  
JOURNAL Patent: JP 1999004692-A 8 12-JAN-1999;  
COMMENT NIPPON SUISAN KAISHA LTD  
OS Unidentified  
PN JP 1999004692-A/8  
PD 12-JAN-1999  
PF 29-SEP-1997 JP 1997264058  
PR

PI SHOICHI YAMADA, KASUTORI VENKATESWARAN,EIJI OHASHI PC  
C12N15/09,C12Q1/68//((C12N15/09,C12R1:07),(C12N15/00,C12N15/00,PC C12R1:07))  
CC Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..21  
FT /organism='Unidentified'.

FEATURES  
source Location/Qualifiers  
1..21  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 9 a 3 c 3 g 6 t  
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
|||||||  
Db 13 TAATATTAC 5

RESULT 8

```

E23445/c      E23445      21 bp      DNA      PAT      07-FEB-2001
LOCUS          Oligonucleotide for detecting Bacillus anthracis and method of
DEFINITION     detection thereof.
E23445
ACCESSION      E23445.1 GI:13024421
VERSION        JP 1999004693-A/8.
KEYWORDS       unidentified.
SOURCE         unidentified.
ORGANISM       unclassified.
REFERENCE      1 (bases 1 to 21)
AUTHORS        Shoichi,Y.K.V.V. and Ohashi.
TITLE          Oligonucleotide for detecting Bacillus anthracis and method of
JOURNAL        detection thereof.
COMMENT        Patent: JP 1999004693-A 8 12-JAN-1999;
              NIPPON SUTSAN KAISHA LTD
              PN JP 1999004693-A/8
              PD 12-JAN-1999
              PF 29-SEP-1997 JP 1997264059
              PR
              PI SHOICHI YAMADA,KASUTORI VENKATESWARAN,EIJI OHASHI PC
              C12N15/09,C12Q1/04,C12Q1/68//((C12N15/09,C12R1:07),(C12Q1/04,PC
              C12R1:07)),
              PC C12N15/00,(C12N15/00,C12R1:07)
              CC Strandedness: Single;
              CC Topology: Linear;
              FH Key Location/Qualifiers
              FT source 1..21 /organism='Unidentified'.
              FT Location/Qualifiers
              source 1..21 /organism='unidentified'
              /db_xref='taxon:32644' 6 t
BASE COUNT     9 a 3 c 3 g 0
ORIGIN
Query Match    100.0%; Score 9; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
Db 13 TAATATTAC 5

RESULT 9
LOCUS          I76530      21 bp      DNA      PAT      03-APR-1998
DEFINITION     Sequence 10 from patent US 5691461.
ACCESSION      I76530
VERSION        I76530.1 GI:3012684
KEYWORDS       Unknown.
SOURCE         Unknown.
ORGANISM       Unclassified.
REFERENCE      1 (bases 1 to 21)
AUTHORS        Ecker,D.J. and Hoke,G.D.
TITLE          Oligonucleotides inhibiting candida germ tube formation
JOURNAL        Patent: US 5691461-A 10 25-NOV-1997;
              Location/Qualifiers
              source 1..21 /organism='unknown'
              source 8 a 4 c 2 g 7 t
BASE COUNT     8 a 4 c 2 g 7 t
ORIGIN
Query Match    100.0%; Score 9; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
Db 13 TAATATTAC 5

E26401/c      E26401      24 bp      DNA      PAT      07-FEB-2001
LOCUS          h-Hyd Protein, polynucleotide encoding said protein, antisense
DEFINITION     polynucleotide thereof and antibody recognizing said protein.
E26401
ACCESSION      E26401.1 GI:13026186
VERSION        JP 1999146786-A/2.
KEYWORDS       unidentified.
SOURCE         unidentified.
ORGANISM       unclassified.
REFERENCE      1 (bases 1 to 24)
AUTHORS        Naoki,N.K.N.N., Nakao,Y.H.M.M. and Hideyuki,S.
TITLE          h-Hyd Protein, polynucleotide encoding said protein, antisense
JOURNAL        polynucleotide thereof and antibody recognizing said protein.
COMMENT        Patent: JP 1999146786-A 2 02-JUN-1999;
              SUMITOMO ELECTRIC IND LTD
              PN JP 1999146786-A/2
              PD 02-JUN-1999
              PF 17-NOV-1997 JP 1997314935
              PR
              PI NAOKI NIBARA,KYOKO NAGAMINE,MITSUYOSHI NAKAO,YOSHIO MI HONDA,
              PI MASAYUKI ANDO,
              PI HIDEYUKI SAYA,
              PC C12N15/09,C07K14/47,C07K16/18,C12P21/02//((C12N15/09,C12R1:91),
              PC (C12P21/02,C12R1:19),C12N15/00,(C12N15/00,C12R1:91) CC
              CC Strandedness: Single;
              CC Topology: Linear;
              FH Key Location/Qualifiers
              FT source 1..24 /organism='Unidentified'.
              FT Location/Qualifiers
              source 1..24 /organism='unidentified'
              /db_xref='taxon:32644' 6 t
BASE COUNT     11 a 5 c 2 g 0
ORIGIN
Query Match    100.0%; Score 9; DB 10; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
Db 4 TAATATTAC 12

RESULT 11
E26401/c      E26401      24 bp      DNA      PAT      07-FEB-2001
LOCUS          h-Hyd Protein, polynucleotide encoding said protein, antisense
DEFINITION     polynucleotide thereof and antibody recognizing said protein.
E26401
ACCESSION      E26401.1 GI:13026186
VERSION        JP 1999146786-A/2.
KEYWORDS       unidentified.
SOURCE         unidentified.
ORGANISM       unclassified.
REFERENCE      1 (bases 1 to 24)
AUTHORS        Naoki,N.K.N.N., Nakao,Y.H.M.M. and Hideyuki,S.
TITLE          h-Hyd Protein, polynucleotide encoding said protein, antisense
JOURNAL        polynucleotide thereof and antibody recognizing said protein.
COMMENT        Patent: JP 1999146786-A 2 02-JUN-1999;
              SUMITOMO ELECTRIC IND LTD
              PN JP 1999146786-A/2
              PD 02-JUN-1999

```

```

PF 17-NOV-1997 JP 1997314935
PR NAKI NIBARA,KYOKO NAGAMINE,MITSUYOSHI NAKAO,YOSHIOMI HONDA,
PI MASAYUKI ANDO,
PI HIDEYUKI SAYA,
PC C12N15/09,C07K14/47,C07K16/18,C12P21/02/(C12N15/09,C12R1:91),
PC (C12P21/02,C12R1:19),C12N15/00,(C12N15/00,C12R1:91) CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..24
FT /organism='Unidentified'.
FEATURES
    source
        1..24
            /organism='unidentified'
            /db_xref='taxon:32644'
BASE COUNT 11 a 5 c 2 g 6 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 11 TAATATTAC 3

RESULT 12
LOCUS AX073421 26 bp DNA PAT 25-JAN-2001
DEFINITION Sequence 9 from Patent WO0102540.
ACCESSION AX073421
VERSION AX073421.1 GI:12583527
KEYWORDS
    synthetic construct.
    synthetic construct
    artificial sequence.
ORGANISM
    1 (bases 1 to 26)
    Hermiston,T., Hawkins,L.K. and Johnson,L.
    Adenoviral vectors for treating disease
    Patent: WO 0102540-A 9 11-JAN-2001;
    ONYX PHARMACEUTICALS, INC. (US)
FEATURES
    Location/Qualifiers
    1..26
        /organism='synthetic construct'
        /db_xref='taxon:32630'
        /note='Oligonucleotide used to mutagenize the E3 region of
        Adenovirus 5.'
BASE COUNT 9 a 4 c 6 g 7 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 10 TAATATTAC 18

RESULT 13
LOCUS AX027097 27 bp DNA PAT 16-SEP-2000
DEFINITION Sequence 12 from Patent DE19855469.
ACCESSION AX027097
VERSION AX027097.1 GI:10188107
KEYWORDS
    synthetic construct.
    artificial sequence.
ORGANISM

```

```

REFERENCE
AUTHORS Esrich,M.
JOURNAL Patent: DE 19855469-A 12 15-JUN-2000;
ESRICH MICHAEL (DE)
FEATURES
    source
        1..27
            /organism='synthetic construct'
            /db_xref='taxon:32630'
            /note='Primer_bind'
BASE COUNT 7 a 7 c 6 g 7 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 5 TAATATTAC 13

RESULT 14
LOCUS MTHSNDU7M 30 bp DNA PRI 22-SEP-1995
DEFINITION H.sapiens mitochondrial NDUFV2 gene intron 7-.
ACCESSION X84435
VERSION X84435.1 GI:995673
KEYWORDS
    NADH ubiquinone oxidoreductase; NDUFV2 gene.
    human.
ORGANISM
    Mitochondrion Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 30)
    De Coe,R.I.F.M.
    Direct Submission
    Submitted (03-FEB-1995) R.I.F.M. De Coe, Human Genetics, University
    Hospital Nijmegen, PO Box 9101, 6500 HB Nijmegen, NETHERLANDS
    2 (bases 1 to 30)
    de Coe,R., Buddiger,P., Smeets,H., Geurts van Kessel,A.,
    Morgan-Hughes,J., Weghuis,D.O., Overhauser,J. and van Oost,B.
    Molecular cloning and characterization of the active human
    mitochondrial NADH:ubiquinone oxidoreductase 24-kDa gene (NDUFV2)
    and its pseudogene
    Genomics 26 (3), 461-466 (1995)
    95331780
    COMMENT
    Sequence overlapping with that under the acc#M22538.
FEATURES
    Location/Qualifiers
    1..30
        /organism='Homo sapiens'
        /organelle='mitochondrion'
        /db_xref='taxon:9606'
        /chromosome='18'
        /clone_lib='The chromosome specific human genome cosmid'
        /clone='120H-11'
        /map='p11.2-p11.31'
        1..30
            /gene='NDUFV2'
            /gene='NDUFV2'
            /note='(-): gene for NADH-ubiquinone oxidoreductase'
            /number=7
            /evidence=experimental
BASE COUNT 9 a 3 c 2 g 16 t
ORIGIN

Query Match 100.0%; Score 9; DB 97; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 11 TAATATTAC 13

```

Db 10 TAATATTAC 18

RESULT 15  
A49364/c 34 bp DNA PAT 07-MAR-1997  
LOCUS A49364 Sequence 45 from Patent WO9607745.  
DEFINITION A49364  
ACCESSION A49364  
VERSION A49364.1 GI:2302843  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.  
REFERENCE 1 (bases 1 to 34)  
AUTHORS Gicquel,B.; Lim,E.M., Portnoi,D., Berthet,F. and Timm,J.  
TITLE MYCOBACTERIA FUNCTIONAL SCREENING AND/OR EXPRESSION VECTORS  
JOURNAL Patent: WO 9607745-A 45 14-MAR-1996;  
PASTEUR INSTITUT (FR)  
COMMENT Other publication FR 2724183 960308.  
FEATURES  
source  
1..34  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 12 a 2 c 3 g 17 t  
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 34;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
|||||  
Db 16 TAATATTAC 8

Search completed: October 24, 2001, 11:42:55  
Job time: 6360 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:49 ; Search time 5701.85 Seconds  
(without alignments)  
84.096 Million cell updates/sec

Title: US-09-462-955-2  
Perfect score: 31  
Sequence: 1 agccgcgggggtaataactagccccgcggct 31

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_bal:\*
- 2: gb\_bal:\*
- 3: gb\_bal:\*
- 4: gb\_bal:\*
- 5: gb\_bal:\*
- 6: gb\_bal:\*
- 7: gb\_bal:\*
- 8: gb\_bal:\*
- 9: gb\_bal:\*
- 10: gb\_bal:\*
- 11: gb\_bal:\*
- 12: gb\_bal:\*
- 13: gb\_bal:\*
- 14: gb\_bal:\*
- 15: gb\_bal:\*
- 16: gb\_bal:\*
- 17: gb\_bal:\*
- 18: gb\_bal:\*
- 19: gb\_bal:\*
- 20: gb\_bal:\*
- 21: gb\_bal:\*
- 22: gb\_bal:\*
- 23: gb\_bal:\*
- 24: gb\_bal:\*
- 25: gb\_bal:\*
- 26: gb\_bal:\*
- 27: gb\_bal:\*
- 28: gb\_bal:\*
- 29: gb\_bal:\*
- 30: gb\_bal:\*
- 31: gb\_bal:\*
- 32: gb\_bal:\*
- 33: gb\_bal:\*
- 34: gb\_bal:\*
- 35: gb\_bal:\*
- 36: gb\_bal:\*
- 37: gb\_bal:\*
- 38: gb\_bal:\*
- 39: gb\_bal:\*
- 40: gb\_bal:\*
- 41: gb\_bal:\*
- 42: gb\_bal:\*
- 43: gb\_bal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------|-------------|
| C 1        | 31    | 100.0   | 1291         | 58    | CFDCG       |
| C 2        | 21.2  | 68.4    | 9294         | 1     | AE006200    |
| C 3        | 20.6  | 66.5    | 981          | 58    | AB000926    |
| C 4        | 20.6  | 66.5    | 989          | 58    | AB000924    |
| C 5        | 20.6  | 66.5    | 990          | 58    | AB000923    |
| C 6        | 20.6  | 66.5    | 209346       | 70    | AC026373    |
| C 7        | 20.2  | 65.2    | 901          | 1     | AF005765    |
| C 8        | 20.2  | 65.2    | 1122         | 1     | AF004344    |

M29963 Coconut fol  
AE006200 Pasteurel  
AB000926 Milk vetc  
AB000924 Milk vetc  
AB000923 Milk vetc  
AC026373 Homo sapi  
AF005765 Unculture  
AF004344 Unidentif



TITLE Direct Submission  
JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology,  
University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN  
55108, USA

FEATURES Location/Qualifiers

source 1..9294  
/organism="Pasteurella multocida"  
/strain="PM70"  
/db\_xref="taxon:747"  
complement(161..853)  
/gene="PM1626"  
complement(161..853)  
/gene="PM1626"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK03710.1"  
/db\_xref="GI:12722027"  
translation="MKTGLIGGMSPESTVYRLINOLINQAKSGNHSAPLMVMNVE  
FEQICLQGEHKAIEILAAKKLTOIGAEGILLATNTMHWKVAPOI IDVIDVPFL  
HIDAYAKAIKINLTOVALLGTFQTMQPEYRDILLHHGITPLVPNGKQOQEHRIIL  
FTELVCYQIKPESQFYLDVINDLAKQAGGVILSGTEIGLLINQODATLFFDFDTTL  
HAQAAAEFLTE"  
complement(872..1102)  
/gene="ilvM"  
/note="PM1627"  
complement(872..1102)  
/gene="ilvM"  
/codon\_start=1  
/transl\_table=11  
/product="ilvM"  
/protein\_id="AAK03711.1"  
/db\_xref="GI:12722028"  
translation="MOTYQFTLDVNRQPTLEIRLVRHRSFYVAMNMTAQEQAN  
IDTVKSDRAFAALLTNOLSKVVDVLAHQVRA"  
complement(1140..2789)  
/gene="ilvG"  
/note="PM1628"  
complement(1140..2789)  
/gene="ilvG"  
/codon\_start=1  
/transl\_table=11  
/product="ilvG"  
/protein\_id="AAK03712.1"  
/db\_xref="GI:12722029"  
translation="MNGARLVTSKTHGVTVTFYGPAGIMPVVDALYDAGTEHLIC  
RNEQGAAMAIYARSGKVGVCATSPGAANLITGLDALMDSIPVVAITQVAP  
LIGTDAFOEDVLGSLACTKHSFLVOSIELPVDVIAKAEIASGRPGVLDIPRD  
IQIAPDTPPTYSVAKPTALSQARHLDTALLKQAKRPVLYVGGVGMQAVPALRE  
FLAVTQMSVSTLKLGSVHPDNPYMGIMGHGTAKAAYAVQECDLLLAFARFDDR  
VTGKLDTAPHAKEVIAHADIDMAETGKLRPPDVALCGDIVQAFNALSIPLIAEWOAHI  
KOLKQTHDFCYSTNOGDTFINPLMLNSLSKKPQSAVITTDVGHQOMWSAQHMOHYA  
PENYITSAGYTGMGFLPAAIGAKKARPDDVEILISGDGSIIMNIOELGTFPKRGKTPV  
KTVLLDNORLGMVQWQBELFNARFNTIILDDNDFVYMLAAAFGIQGERITKGEVEEG  
ADRLLSAEAFVFLHVCISSNNVWLPVPNACNLDNWEQM"  
3149..3967  
/gene="PM1629"  
3149..3967  
/gene="PM1629"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK03713.1"  
/db\_xref="GI:12722030"  
translation="MSFPFRAIVSDLDGTLLNAHMGIDGTITQTLQOLAAKIDIML  
ATGRNHTDLLPILKKVNIKAVMITNSGARAQDLQGNLLVROYLPQIAFDIMNLDFF  
RQRCVSAYQGDWFIINIDVQPKYHQDSGFMYEVDFQAHHGRETERKVFFIGREPO  
DLGLEHLOANADTTSITYIPVCLFINKNKVSASALEKVLADRDYDLQCLAFG  
DGMNDVQWLSRGVCGVGNADPRLKAEACHPLEIVGNNAQESVANYIRTVFDE"  
3983..4387  
/gene="crcB"  
/note="PM1630"  
3983..4387

gene  
/gene="crcB"  
/codon\_start=1  
/transl\_table=11  
/product="crcB"  
/protein\_id="AAK03714.1"  
/db\_xref="GI:12722031"  
translation="MAGSRSVISVGQOIIIFISSGAALGALSRWGLGLLLNPLFSAFSL  
GTILVYGLGLIIGVFLAFWVQPCSAEMRLFFVFGLSLTITSTFSAEVIENLIO  
QKWLAGLMLASGHLGLCLLTALGVFIWYQW"  
complement(4365..5318)  
/gene="pldB"  
/note="PM1631"  
complement(4365..5318)  
/gene="pldB"  
/codon\_start=1  
/transl\_table=11  
/product="pldB"  
/protein\_id="AAK03715.1"  
/db\_xref="GI:12722032"  
translation="MIREPHFTFALAEHPFAQFPLOQDVQGGKGVRTYRHFVQDN  
PVORNLVILVNGRAENLKWTELAVDFYQOQYDVLFVDRGQGYSORLLPDHKGYYID  
EPFYADDMAILLAKISALYPVENOHILAHLSIGALISTYYLANYDHQVKSAPVSAFY  
GIPKHSFRDELINLMLLGGQSRVYVFGKGRIPADLDNLSCCRTRMWMNRINR  
RYPHILGGPFRWHLCESAIKLPTILPRIEIPVILLOSEKKEKIVENKKNLRLTAL  
LPREUQIAQSKHEILFPERDLIRSEALKKISQFFTANSK"  
complement(5460..6569)  
/gene="asd"  
/note="PM1632"  
complement(5460..6569)  
/gene="asd"  
/codon\_start=1  
/transl\_table=11  
/product="asd"  
/protein\_id="AAK03716.1"  
/db\_xref="GI:12722033"  
translation="MONVGFIGWRMGVSVLMDRMOENDEANINPVFTTSQAGQNA  
PVEGKQAGELKDAFDIAELQKLDIIVTCQGGDYTVNEYPKLRAGKMGVWIDAASAL  
RMEKDAIIVDPNONVNSEGLKNGIKTYNGNCTVSLMLMALGGLFERDLVEMVSVA  
TYAASGAGAKNMRELLVOMGOLEDVSKVDLNPASSILDIERKVTAKMGDLP IENF  
GAPLAGSLIPWIDKLWEDGOTKEWKGVAETNKILGDDSP I PVDGLCVRIGALRCHS  
QAFITMKKDLPLLEEIOILASHNEWKVI I NEKDVTLRELTPKVTGTLSVPVGRLR  
KLKMGPEYLAAPTVGDLQLLWGAAEPLRLILKQLVA"  
6770..7009  
/gene="PM1633"  
6770..7009  
/gene="PM1633"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK03717.1"  
/db\_xref="GI:12722034"  
translation="MDFLCFVAFVPRDERYQKRSYVVIQPTERKLGKASISURNFSD  
SDTESKAWTIDINIKIPEENKIFNKKSTELKFPQ"  
7028..7381  
/gene="PM1634"  
7028..7381  
/gene="PM1634"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK03718.1"  
/db\_xref="GI:12722035"  
translation="MNPNMKKKILEKSKGLPLFALLAEFEDYFGSVKDSDETKELFLS  
FVAEIMHDGELKFAIRKGFLEGSTEEQIDLFROAWPDHYDDEKLEVIDIHLWITVAP  
AGAYWICEDYEWT"  
7583..7948  
/gene="PM1635"  
7583..7948  
/gene="PM1635"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK03719.1"

Thu Oct 25 13:08:39 2001

```

/db_xref="GI:12722036"
/translation="MLEKSEAQQYKRIVENRGLVLDGLFANFQCYGDIRDKPEIKE
MFLMLADLMKTGELKATKGFLEGSIEEQIDLFQAPWDHYDDKELEYDIDLHWI
TYAPAGAVWICDGYEDWI"
8272..8637
/gene="PM1636"
CDS
8272..8637
/gene="PM1636"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03720.1"
/db_xref="GI:12722037"
/translation="MLEKSEAQQYKRIVENRGLTLIGMFSFQVCYGDIRDKPEIKE
MFLMLADLMKTGELKATKGFLEGSIEEQIDLFQAPWDHYDDKELEYDIDLHWI
VYAPAGAVWICDGYEDWT"
8707..9144
/gene="PM1637"
CDS
8707..9144
/gene="PM1637"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03721.1"
/db_xref="GI:12722038"
/translation="MMDYRRWSSRPDERYQRDKVNMLEKSEAQQYKRVVNNRGL
ALDGLFASQVCGDIQDKPEIKEMFELMADLMKTGELKATKGFLEGSIEEQIDV
FQAPWDHYDDKELEYDIDLHWIVYAPAGAVWICDGYEDWT"
BASE COUNT 2772 a 1733 c 1979 g 2810 t
ORIGIN

Query Match 68.4%; Score 21.2; DB 1; Length 9294;
Best Local Similarity 88.5%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gcggggtaatactagcccgagcgc 30
||||| ||||||| |||||
Db 676 GCGGGGAATACTACCCACTCGGC 651

RESULT 3
AB000926/c 981 bp DNA circular VRL 30-OCT-1998
LOCUS
DEFINITION Milk vetch dwarf virus genome segment 7 encoding viral
non-structural protein, complete sequence.
ACCESSION AB000926
VERSION AB000926.1 GI:3808185
KEYWORDS viral non-structural protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM
Viruses; ssDNA viruses; Nanovirus.
REFERENCE
1 (bases 1 to 981)
Sano,Y.
Direct Submission
Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel:075-724-7764, Fax:075-724-7764)
2 (sites)
Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
99094638
MEDLINE
JOURNAL
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798663.
FEATURES
source
1..981
/organism="milk vetch dwarf virus"
/virion
/db_xref="taxon:67585"
/chromosome="5"
/lab_host="Pisum sativum"
6..34
stem_loop 279..285
TATA_signal 376..828
CDS
/function="unknown"
/notes="putative"
/codon_start=1
/product="viral non-structural protein"
/protein_id="BAA33984.1"
/db_xref="GI:3798660"
/translation="MSVNDYSLFDEATDELVSERKLIADVCHDDDDSOVINVKVEDIF
CDMSDKVILQFRLCYKRYKRLDITLLGCRMKVYTELTPSERSLKSILOKMNIIIS
DGNLYILRMFFINQLINCKWITREDVPICTLYHVNNTPIVDI"
BASE COUNT 284 a 167 c 210 g 328 t

```

```

stem_loop 6..34
TATA_signal 280..286
CDS 346..723
/function="unknown"
/notes="putative"
/codon_start=1
/product="viral non-structural protein"
/protein_id="BAA33986.1"
/db_xref="GI:3798664"
/translation="MEDFKQPKLSYGEIVOMKEODAFWSCYHFEFLRNEDVIGEMCR
RHGRKLPAKPLPTAPIRHWLKTALIDVRVDECKSCSHEISRRDYNPIKKEGLKD
LYDSGNRYQVYSSNCRKSD"
BASE COUNT 291 a 170 c 214 g 306 t
ORIGIN

Query Match 66.5%; Score 20.6; DB 58; Length 981;
Best Local Similarity 85.2%; Pred. No. 4.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 gcggggtaatactagcccgagcgc 31
||||| ||||||| |||||
Db 31 GCGGGGTAATAAGCCCGGCCCT 5

RESULT 4
AB000924/c 989 bp DNA circular VRL 30-OCT-1998
LOCUS
DEFINITION Milk vetch dwarf virus genome segment 5 encoding viral
non-structural protein, complete sequence.
ACCESSION AB000924
VERSION AB000924.1 GI:3808183
KEYWORDS viral non-structural protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM
Viruses; ssDNA viruses; Nanovirus.
REFERENCE
1 (bases 1 to 989)
Sano,Y.
Direct Submission
Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel:075-724-7764, Fax:075-724-7764)
2 (sites)
Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
99094638
MEDLINE
JOURNAL
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798659.
FEATURES
source
1..989
/organism="milk vetch dwarf virus"
/virion
/db_xref="taxon:67585"
/chromosome="5"
/lab_host="Pisum sativum"
6..34
stem_loop 279..285
TATA_signal 376..828
CDS
/function="unknown"
/notes="putative"
/codon_start=1
/product="viral non-structural protein"
/protein_id="BAA33984.1"
/db_xref="GI:3798660"
/translation="MSVNDYSLFDEATDELVSERKLIADVCHDDDDSOVINVKVEDIF
CDMSDKVILQFRLCYKRYKRLDITLLGCRMKVYTELTPSERSLKSILOKMNIIIS
DGNLYILRMFFINQLINCKWITREDVPICTLYHVNNTPIVDI"
BASE COUNT 284 a 167 c 210 g 328 t

```





Consensus quality: 201501 bases at least Q20  
Estimated insert size: 193981; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 41 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 19248: contig of 19248 bp in length  
19249 19348: gap of unknown length  
19349 40153: contig of 20805 bp in length  
40154 40253: gap of unknown length  
40254 52177: contig of 11924 bp in length  
52178 52277: gap of unknown length  
52278 64996: contig of 12719 bp in length  
64997 65096: gap of unknown length  
65097 77175: contig of 12079 bp in length  
77176 77275: gap of unknown length  
77276 85780: contig of 8505 bp in length  
85781 85880: gap of unknown length  
85881 94536: contig of 8656 bp in length  
94537 94636: gap of unknown length  
94637 100914: contig of 6278 bp in length  
100915 101014: gap of unknown length  
101015 107274: contig of 6260 bp in length  
107275 107374: gap of unknown length  
107375 114125: contig of 6751 bp in length  
114126 114225: gap of unknown length  
114226 118662: contig of 4437 bp in length  
118663 118762: gap of unknown length  
118763 126267: contig of 7505 bp in length  
126268 126367: gap of unknown length  
126368 132591: contig of 6224 bp in length  
132592 132691: gap of unknown length  
132692 140866: contig of 8175 bp in length  
140867 140966: gap of unknown length  
140967 147869: contig of 6903 bp in length  
147870 147969: gap of unknown length  
147970 153425: contig of 5456 bp in length  
153426 153525: gap of unknown length  
153526 158844: contig of 5319 bp in length  
158845 158944: gap of unknown length  
158945 163397: contig of 4453 bp in length  
163398 163497: gap of unknown length  
163498 167168: contig of 3671 bp in length  
167169 167268: gap of unknown length  
167269 170619: contig of 3351 bp in length  
170620 170719: gap of unknown length  
170720 174986: contig of 4267 bp in length  
174987 175086: gap of unknown length  
175087 177726: contig of 2640 bp in length  
177727 177826: gap of unknown length  
177827 180359: contig of 2533 bp in length  
180360 180459: gap of unknown length  
180460 182486: contig of 2027 bp in length  
182487 182586: gap of unknown length  
182587 184445: contig of 1859 bp in length  
184446 184545: gap of unknown length  
184546 186654: contig of 2109 bp in length  
186655 186754: gap of unknown length  
186755 188617: contig of 1863 bp in length  
188618 188717: gap of unknown length  
188719 191290: contig of 2573 bp in length  
191291 191390: gap of unknown length  
191391 193015: contig of 1625 bp in length  
193016 193115: gap of unknown length

193116 194288: contig of 1173 bp in length  
194289 194388: gap of unknown length  
194389 195747: contig of 1359 bp in length  
195748 195847: gap of unknown length  
195848 197563: contig of 1716 bp in length  
197564 197663: gap of unknown length  
197664 198798: contig of 1135 bp in length  
198799 198898: gap of unknown length  
198900 200522: contig of 1624 bp in length  
200523 200622: gap of unknown length  
200623 201905: contig of 1283 bp in length  
201906 202005: gap of unknown length  
202006 203174: contig of 1169 bp in length  
203175 203274: gap of unknown length  
203275 204491: contig of 1217 bp in length  
204492 204591: gap of unknown length  
204592 205794: contig of 1203 bp in length  
205795 205894: gap of unknown length  
205896 207005: contig of 1111 bp in length  
207006 207105: gap of unknown length  
207106 208212: contig of 1107 bp in length  
208213 208312: gap of unknown length  
208313 209346: contig of 1034 bp in length.

FEATURES  
source  
1. 209346  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-353N14"

BASE COUNT 46917 a 55642 c 55603 g 47116 t 4068 others  
ORIGIN

Query Match 66.5%; Score 20.6; DB 70; Length 209346;  
Best Local Similarity 85.2%; Pred. No. 94;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgc 27  
|||||  
Db 25724 AGCCGCGGGGTAGTCTAGCCCGCAGC 25750

RESULT 7

AF005765/c AF005765 901 bp DNA BCT 23-NOV-1998  
LOCUS Uncultured archaeon Arc.171 16S ribosomal RNA gene, partial  
DEFINITION sequence.

ACCESSION AF005765.1 GI:2443507

VERSION AF005765

KEYWORDS uncultured archaeon Arc.171.

SOURCE uncultured archaeon Arc.171

ORGANISM Archaea; environmental samples.

REFERENCE 1 (bases 1 to 901)

AUTHORS Chandler, D.P., Brockman, F.J., Bailey, T.J. and Fredrickson, J.K.

TITLE Phylogenetic diversity of Archaea and Bacteria in a deep subsurface

JOURNAL Microb. Ecol. 36, 37-50 (1998)

REFERENCE 2 (bases 1 to 901)

AUTHORS Chandler, D.P.

TITLE Direct Submission

JOURNAL Submitted (29-MAY-1997) Environmental Microbiology, Battelle

Pacific Northwest National Lab, 900 Battelle Blvd., Mail Stop

K4-06, Richland, WA 99352, USA

FEATURES Location/Qualifiers

source 1. .901

/organism="uncultured archaeon Arc.171"

/db\_xref="taxon:62249"

/clone="Arc.171"

<1. .901

/product="16S ribosomal RNA"

BASE COUNT 169 a 296 c 215 g 221 t

ORIGIN

```

Query Match      65.2%; Score 20.2; DB 1; Length 901;
Best Local Similarity 88.0%; Pred. No. 6.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagccccc 25
    ||||| | ||||| | ||||| | ||||| |
Db 420 AGCCGCGGGTAATACCGCCCC 396

RESULT 8
AF004344
LOCUS AF004344 1122 bp DNA BCT 25-MAR-1998
DEFINITION Unidentified crenarchaeote 16S ribosomal RNA gene, partial
sequence.
ACCESSION AF004344
VERSION AF004344.1 GI:2984608
KEYWORDS .
SOURCE uncultured crenarchaeote.
ORGANISM Archaea; Crenarchaeota; environmental samples.
REFERENCE 1 (bases 1 to 1122)
AUTHORS Vetrilani,C., Reysenbach,A.L. and Dore,J.
TITLE Recovery and phylogenetic analysis of archaeal rRNA sequences from
continental shelf sediments
JOURNAL FEMS Microbiol. Lett. 161 (1), 83-88 (1998)
MEDLINE 98222521
REFERENCE 2 (bases 1 to 1122)
AUTHORS Vetrilani,C., Reysenbach,A.-L. and Dore,J.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-1997) Center of Marine Biotechnology, Maryland
Biotechnology Institute, 701 E. Pratt St., Suite 236, Baltimore, MD
21202, USA

FEATURES
source
Location/Qualifiers
1. .1122
/organism="uncultured crenarchaeote"
/db_xref="taxon:29281"
/clone="BBA4"
/notes="Isolated from Buzzards Bay sediments, NW Atlantic
Ocean, depth 12 m
type: Clone BBA4"
<1. >1122
/product="16S ribosomal RNA"
BASE COUNT 273 a 273 c 367 g 209 t
ORIGIN

rRNA

Query Match      65.2%; Score 20.2; DB 1; Length 1122;
Best Local Similarity 88.0%; Pred. No. 6.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagccccc 25
    ||||| | ||||| | ||||| | ||||| |
Db 134 AGCCGCGGGTAATACCGCCCC 158

RESULT 9
AY016473
LOCUS AY016473 804 bp DNA BCT 11-JAN-2001
DEFINITION Uncultured Front Range soil crenarchaeote FRA31B 16S ribosomal RNA
gene, partial sequence.
ACCESSION AY016473
VERSION AY016473.1 GI:12083488
KEYWORDS .
SOURCE uncultured Front Range soil crenarchaeote FRA31B.
ORGANISM Archaea; Crenarchaeota; environmental samples.
REFERENCE 1 (bases 1 to 804)
AUTHORS Oline,D.K., Schmidt,S.K. and Grant,M.C.
TITLE Diversity and Distribution of Soil Crenarchaeota in Coniferous
Forest and Alpine Tundra Soils in the Front Range of Colorado
Unpublished

```

```

REFERENCE 2 (bases 1 to 804)
AUTHORS Oline,D.K., Schmidt,S.K. and Grant,M.C.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2000) Environmental, Population, and Organismic
Biology, University of Colorado at Boulder, Campus Box 334,
University of Colorado, Boulder, CO 80309, USA
COMMENT Ponderosa pine forest soil isolate, USA: Front Range of Colorado,
Al site, elev. 1800m.m.
FEATURES
source
Location/Qualifiers
1. .804
/organism="uncultured Front Range soil crenarchaeote
FRA31B"
/isolate="FRA31B"
/db_xref="taxon:147503"
/notes="Ponderosa pine forest soil isolate
?: USA: Front Range of Colorado, Al site, elev. 1800m."
<1. >804
/product="16S ribosomal RNA"
BASE COUNT 205 a 188 c 254 g 156 t
ORIGIN

Query Match      64.5%; Score 20; DB 2; Length 804;
Best Local Similarity 82.1%; Pred. No. 8.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccg 28
    ||||| | ||||| | ||||| | ||||| |
Db 431 AGCCGCGGGTAATACCGACCGGAG 458

RESULT 10
AF112330
LOCUS AF112330 519 bp DNA VRT 26-FEB-2001
DEFINITION Galaxias paucispindylus 16S ribosomal RNA gene, mitochondrial gene
for mitochondrial RNA, partial sequence.
ACCESSION AF112330
VERSION AF112330.1 GI:7688534
KEYWORDS Galaxias paucispindylus.
SOURCE Mitochondrion Galaxias paucispindylus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Osmeriformes; Galaxiidae; Galaxias.
REFERENCE 1 (bases 1 to 519)
AUTHORS Waters,J.M., Lopez,J.A. and Wallis,G.P.
TITLE Molecular phylogenetics and biogeography of galaxiid fishes
(Osteichthyes: Galaxiidae): dispersal, vicariance and the position
of Lepidogalaxias salmandroides
JOURNAL Syst. Biol. 49 (4), 777-795 (2000)
REFERENCE 2 (bases 1 to 519)
AUTHORS Waters,J.M. and Wallis,G.P.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1998) Department of Zoology, University of Otago,
PO Box 56, Dunedin, Otago, New Zealand
FEATURES
source
Location/Qualifiers
1. .519
/organism="Galaxias paucispindylus"
/organelle="mitochondrion"
/db_xref="taxon:89560"
<1. >519
/product="16S ribosomal RNA"
BASE COUNT 144 a 126 c 129 g 120 t
ORIGIN

rRNA

Query Match      63.9%; Score 19.8; DB 8; Length 519;
Best Local Similarity 91.3%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ccgcgggggtaatactagccccc 25
    ||||| | ||||| | ||||| | ||||| |
Db 288 CGCGGGGAAAGACTAGCCCC 310

```

| FEATURES   | source | Location/Qualifiers   |
|------------|--------|---|
| Qy         | 3      | ccgcgggggtaatactagccccc 25  |
| Db         | 288    | CCGCGGGGGAAGACTAGCCCC 310   |
| LOCUS      |        | AF022100.1 GI:2502045   |
| DEFINITION |        | Galaxias truttaceus.  |
| ACCESSION  |        | AF022100  |
| VERSION    |        | 1   |
| KEYWORDS   |        | Galaxias truttaceus.  |
| ORGANISM   |        | Mitochondrion Galaxias truttaceus   |
| REFERENCE  |        | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Osmeriformes; Galaxiidae; Galaxias. |
| AUTHORS    |        | Waters, J.M., and White, R.W.G.   |
| TITLE      |        | Molecular phylogeny and biogeography of the Tasmanian and New Zealand mudfishes (Salmoniformes: Galaxiidae)   |
| JOURNAL    |        | Aust. J. Zool. 45, 39-48 (1997)   |
| AUTHORS    |        | 2 (bases 1 to 521)  |
| TITLE      |        | 2 (bases 1 to 521)  |
| JOURNAL    |        | Submitted (02-SEP-1997) Biology, Virginia Commonwealth University, Park Ave, Richmond, VA 23284, USA  |
| DEFINITION |        | Location/Qualifiers   |
| ACCESSION  |        | 1. .521   |
| VERSION    |        | /organism="Galaxias truttaceus"   |
| KEYWORDS   |        | /organism="Galaxias truttaceus"   |
| ORGANISM   |        | /db_xref="taxon:66448"  |
| REFERENCE  |        | <1..>521  |
| AUTHORS    |        | /product="16S ribosomal RNA"  |
| TITLE      |        | 137 a 133 c 133 g 118 t   |
| JOURNAL    |        | BASE COUNT  |
| DEFINITION |        | ORIGIN  |
| ACCESSION  |        | Query Match   |
| VERSION    |        | Best Local Similarity   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        | Matches   |
| ORGANISM   |        | Matches   |
| REFERENCE  |        | Matches   |
| AUTHORS    |        | Matches   |
| TITLE      |        | Matches   |
| JOURNAL    |        | Matches   |
| DEFINITION |        | Matches   |
| ACCESSION  |        | Matches   |
| VERSION    |        | Matches   |
| KEYWORDS   |        |   |

```

non-structural protein, complete sequence.
ACCESSION AB000925 GI:3808184
VERSION 1
KEYWORDS viral non-structural protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM milk vetch dwarf virus
REFERENCE 1 Viruses: ssDNA viruses; Nanovirus.
1 (bases 1 to 977)
AUTHORS Sano,Y.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel:075-724-7764. Fax:075-724-7764)
2 (sites)
REFERENCE Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
AUTHORS Sequences of ten circular ssDNA components associated with the milk
TITLE vetch dwarf virus genome
JOURNAL J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
MEDLINE 99094638
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798661.
FEATURES
    Location/Qualifiers
        source
            1..977
            /organism="milk vetch dwarf virus"
            /viral
            /db_xref="taxon:67585"
            /chromosome="segment 6"
            /lab_host="Pisum sativum"
            1..33
            304..310
            363..824
            /function="unknown"
            /note="putative"
            /codon_start=1
            /product="viral non-structural protein"
            /protein_id="BAA33985.1"
            /db_xref="GI:3798662"
            /translation="MADWFAFLKCTHVCDFPTLACDPSQEIFCCSMKNKLNDSRK
            VLLVSCVSFGNSFYGNRNVRGOISMLEDDGVCRPIGVPIGGYLYHNDYGYVEG
            EKTENLDIESQYLKDEYDNKRFVSVNLNGLSLCDLKVFIHVHAIKRV"
        polyA_signal 823..828
        BASE COUNT 275 a 162 c 213 g 327 t
        ORIGIN
            63.9%; Score 19.8; DB 58; Length 977;
            Best Local Similarity 91.3%; Pred. No. 9.6e-02;
            Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
            QY 5 gcgggggtaatactagccccgc 27
            Db 28 GCGGGGGTAATAAGCCCCGC 6
            |||||
RESULT 15
LOCUS AB000927/c 985 bp DNA circular VRL 30-OCT-1998
DEFINITION Milk vetch dwarf virus genome segment 8 encoding viral
non-structural protein, complete sequence.
ACCESSION AB000927
VERSION 1
KEYWORDS viral non-structural protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM milk vetch dwarf virus
REFERENCE 1 Viruses: ssDNA viruses; Nanovirus.
1 (bases 1 to 985)
AUTHORS Sano,Y.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel:075-724-7764, Fax:075-724-7764)
2 (sites)

```

```

AUTHORS Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
TITLE Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
JOURNAL J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
MEDLINE 99094638
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798665.
FEATURES
    Location/Qualifiers
        source
            1..985
            /organism="milk vetch dwarf virus"
            /viral
            /db_xref="taxon:67585"
            /chromosome="segment 8"
            /lab_host="Pisum sativum"
            1..33
            276..282
            344..682
            /function="unknown"
            /note="putative"
            /codon_start=1
            /product="viral non-structural protein"
            /protein_id="BAA33987.1"
            /db_xref="GI:3798666"
            /translation="MADPVYOGYODGDIDAOKRHOALYIGIILIMVCIILWVC
            IMLACYIPGFLKKTMEAWLSSSSMKRVAATITRTPEATGPERNWDAROTNAA
            SSQPSNGGVF"
        polyA_signal 775..780
        BASE COUNT 283 a 180 c 213 g 309 t
        ORIGIN
            63.9%; Score 19.8; DB 58; Length 985;
            Best Local Similarity 91.3%; Pred. No. 9.5e-02;
            Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
            QY 5 gcgggggtaatactagccccgc 27
            Db 28 GCGGGGGTAATAAGCCCCGC 6
            |||||
Search completed: October 24, 2001, 11:42:53
Job time: 6:58 sec

```

us-09-462-955-2.rge

Thu Oct 25 13:08:39 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:48 ; Search time 5701.85 Seconds  
(without alignments)  
762.285 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_711\_991  
Perfect score: 281  
Sequence: 1 cacatgtgtggaaccagaa.....ttaagtgatgtcatctaa 281

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

GenEmbl:\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_om.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_pl1.\*
- 13: gb\_pl2.\*
- 14: gb\_pl3.\*
- 15: gb\_pl4.\*
- 16: em\_ba1.\*
- 17: em\_ba2.\*
- 18: em\_fun.\*
- 19: em\_htgo\_hum.\*
- 20: em\_htgo\_inv.\*
- 21: em\_htgo\_rod.\*
- 22: em\_htg\_hum1.\*
- 23: em\_htg\_hum2.\*
- 24: em\_htg\_hum3.\*
- 25: em\_htg\_hum4.\*
- 26: em\_htg\_hum5.\*
- 27: em\_htg\_hum6.\*
- 28: em\_htg\_hum7.\*
- 29: em\_htg\_hum8.\*
- 30: em\_htg\_inv1.\*
- 31: em\_htg\_inv2.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_rod.\*
- 34: em\_hum1.\*
- 35: em\_hum2.\*
- 36: em\_hum3.\*
- 37: em\_hum4.\*
- 38: em\_hum5.\*
- 39: em\_hum6.\*
- 40: em\_hum7.\*
- 41: em\_in.\*
- 42: em\_om.\*
- 43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_rod.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_vil3.\*  
61: gb\_vil4.\*  
62: gb\_vil5.\*  
63: gb\_vil6.\*  
64: gb\_vil7.\*  
65: gb\_vil8.\*  
66: gb\_vil9.\*  
67: gb\_vil10.\*  
68: gb\_vil11.\*  
69: gb\_vil12.\*  
70: gb\_vil13.\*  
71: gb\_vil14.\*  
72: gb\_vil15.\*  
73: gb\_vil16.\*  
74: gb\_vil17.\*  
75: gb\_vil18.\*  
76: gb\_vil19.\*  
77: gb\_vil20.\*  
78: gb\_vil21.\*  
79: gb\_vil22.\*  
80: gb\_vil23.\*  
81: gb\_vil24.\*  
82: gb\_vil25.\*  
83: gb\_vil26.\*  
84: gb\_vil27.\*  
85: gb\_vil28.\*  
86: gb\_vil29.\*  
87: gb\_vil30.\*  
88: gb\_vil31.\*  
89: gb\_vil32.\*  
90: gb\_vil33.\*  
91: gb\_vil34.\*  
92: gb\_vil35.\*  
93: gb\_vil36.\*  
94: gb\_vil37.\*  
95: gb\_vil38.\*  
96: gb\_vil39.\*  
97: gb\_vil40.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description        |
|------------|-------|---------|--------------|-------|--------------------|
| 1          | 281   | 100.0   | 1291         | 58    | CFDCG              |
| 2          | 89.4  | 31.8    | 1007         | 58    | M29963 Coconut fol |
| 3          | 86.2  | 30.7    | 1004         | 59    | AJ132187 faba bean |
| 4          | 74    | 26.3    | 1007         | 58    | AJ005966 faba bean |
| 5          | 70.8  | 25.2    | 1091         | 9     | AB000920 Milk vetc |
| 6          | 70.8  | 25.2    | 1096         | 9     | AR063453 Sequence  |
| 7          | 70.8  | 25.2    | 1096         | 9     | AR063452 Sequence  |
| 8          | 70.8  | 25.2    | 1106         | 58    | BTIV2              |
|            |       |         |              |       | L32167 Banana bunc |
|            |       |         |              |       | AR063451 Sequence  |

```

9 70.8 25.2 1106 58 BTV1
10 70.6 25.1 1095 58 AF216222 Banana bunc
11 70.6 25.1 1111 58 BB012586
12 69.8 24.8 1095 58 BB012586
13 69.2 24.6 1009 58 AB000921
14 69 24.6 1127 58 BB012587
15 68.6 24.4 1004 58 NV5968
16 67.4 24.0 1109 58 AF216221 Banana bu
17 65.2 23.2 1017 59 SCU16735
18 61.8 22.0 1022 58 FBNVREP
19 58.4 20.8 1002 58 AB000922
20 55.6 19.8 1367 58 AY2238493
21 53.6 19.1 1000 58 AB000922
22 53 18.9 1375 58 CLE132345
23 53 18.9 1376 58 CLE132344
24 52.2 18.6 1015 58 FBEL132185
25 49.8 17.7 1014 58 NV5964
26 47.6 16.9 1022 58 AB000947
27 44.6 15.9 1105 9 AR010234
28 44.4 15.8 1103 9 AR010232
29 44.4 15.8 1103 58 AF238874
30 44.4 15.8 1110 9 AR010230
31 43 15.3 1111 9 AR010228
32 43 15.3 1111 58 AF238875
33 42.8 15.2 1104 9 AR010233
34 42.8 15.2 1104 12 ACV14168
35 42.2 15.0 1343 12 ACV14168
36 42.2 15.0 161361 71 AC037465
37 41.4 14.7 982 9 AR112007
38 41.4 14.7 1109 9 AR010229
39 41.4 14.7 1110 58 BB018077
40 41.4 14.7 1111 9 AR010225
41 41.4 14.7 1111 9 AR010240
42 41.4 14.7 1111 58 AF102780
43 41.4 14.7 1111 59 S56276
44 41.2 14.7 1103 58 AF246123
45 40 14.2 1111 9 AR010226

```

# ALIGNMENTS

```

RESULT 1
CFDGC 1291 bp ss-DNA circular VRL
LOCUS Coconut foliar decay virus, complete genome.
DEFINITION M29963
ACCESSION M29963
VERSION 1 GI:323306
KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSUA3A.
ORGANISM
Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1291)
AUTHORS Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
JOURNAL Virology
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by W.Rhode, for release after publication.
FEATURES
source
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40..70
/notes="stem-loop structure"
103..975
/notes="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSIRRCFTLNHYETEFAANVVRRIESLNLYVAIVGDEVAPS
TGQRHLQGFTHLTKRRRLQGLKLVIGNDRIHLEPIRGSDQNRDYSCKSERVLLHGV

```

```

BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN
Query Match 100.0%; Score 281; DB 58; Length 1291;
Best Local Similarity 100.0%; Pred. No. 9.6e-67;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cacatgtggtgaacacagagcgtattgtaccagtcacatcgagcagcacaacgaa 60
DB 711 CACATGTGTGGAAACAGAAAGACGCTATTGTACCAGTACATCGAGGACCAACGAA 770
QY 61 tttaacctcgatgtaccaggtgaatttagagattattaaattatgccctgttaga 120
DB 771 TTTAATCCTCGATGTACCCAGGTGAATTTAGAGTATTTAAATTTATGCCCTGT 830
QY 121 tgttaagaacagggcattcagtcggacaatacagaccccttagttatcttgggtt 180
DB 831 TGTTAAGAACAGGGCATTTCAGTTCGACAAATACCAACCCCTTAGTTATCTGG 890
QY 181 ccatgtgcattactcgtatttggcaattgcctgcctgattatttgaataatcac 240
DB 891 CCAATGTCATGTACTCGTATTTCCTCAATGCTGCTGCTGCTGCTGCTGCTG 950
QY 241 cagaataaaactgtggaatattttaaagtattgtgcatctaa 281
DB 951 CAGATAAAACTGTGGAATATTTAAAGTAGTATTTGAAATATCAGCAGGGA 991
RESULT 2
FBE132187
LOCUS
DEFINITION faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian

```

```

TRPVKRPRLAQRAREPDLRLDPGCGYRRVCVHVHGAASVEWTRWAENPFPFPHNQ
LEVLSAICPEADRTILWICRGDCKSVFAYGLKLPDMFVTCGTRKDVLYQIE
DPKRLILIDVPRCNLEYLNALCEVKNRAFSDDKYEPLSYLGFEDHVVHVLVFNVLDP
YIKISDRIRIKWNI
314..775
/notes="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGETWGRPVVPTNRIETTVRRNGCPSSTESRLVLESKGDWPN
DLRLNLSANKTQADTALYELRWNGQDGLKIRSHPHITIGSLKCCLESORT
IAQSSGYADETEETGSPCLPNISDSSPTGTHVVEPERTYCTSTSTQNEI"
complement(422..568)
/notes="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFORPILSIPPKLRVQIRLGLPGLGGVHVQVPOIVGP
IVAF"
639..797
/notes="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MRTRRRRREVRCQISRTOARLVLMWNQKGRIVPVHRGPKTK
FNPRCTQV"
complement(823..987)
/notes="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHTLNIPQFYSVPADFQIIRQIDGKYEYMHMVEPKITKGFFV
RTECPVLNTP"
1098..1286
/notes="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGPTIKDSIWRNLLCLOCTOPLSTSPQVSSLLEKK
AASLYPSICFCAIGRLS"

```



EV1-93.  
 AJ132187.1 GI:4995171  
 C9-Eg gene; rep protein.  
 faba bean necrotic yellows virus.  
 faba bean necrotic yellows virus  
 Viruses; ssDNA viruses; Nanovirus.  
 1 (bases 1 to 1007)  
 Ten distinct circular ssDNA components, four of which encode  
 putative replication-associated proteins, are associated with the  
 faba bean necrotic yellows virus genome  
 J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)  
 99094637  
 2 (bases 1 to 1007)  
 Timchenko, T., de Kouchkovsky, F., Katul, L., David, C., Vetten, H.J.  
 and Gronenborn, B.  
 A single rep protein initiates replication of multiple genome  
 components of faba bean necrotic yellows virus, a single-stranded  
 DNA virus of plants  
 J. Virol. 73 (12), 10173-10182 (1999)  
 20027244  
 3 (bases 1 to 1007)  
 Katul, L.  
 Direct Submission  
 Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie,  
 Mikrobiol., u. biol. Sicherheit, Biologische Bundesanstalt f. Land-  
 u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY  
 Location/Qualifiers  
 1. .1007  
 /organism="faba bean necrotic yellows virus"  
 /virion  
 /isolate="Egyptian EV1-93"  
 /db\_xref="taxon:59817"  
 /lab\_host="Vicia faba"  
 1. .39  
 73. .1003  
 /gene="C9-Eg"  
 73. .918  
 /gene="C9-Eg"  
 /function="replication-associated protein"  
 /note="32.53 kba"  
 /codon\_start=1  
 /evidence="experimental"  
 /product="Rep protein"  
 /protein\_id="CAB44027.1"  
 /db\_xref="GI:4995172"  
 /db\_xref="SPTREMBL:Q9WIK2"  
 /translation="MSAVNVVFTLNFAGEVPVLSFDERVOYAVWQHRYVNHDIQGV  
 OLKKAKNTVKNIIIGGNPHLEKMGSIIEASAYAOKEESRVAGPWSYGEGLKKGSHK  
 RKIMELIKDPENLEERPKYBRAMANSAMDESKLAEEGGFPYTLTYSWQETVLGLLEE  
 EPNDRTIIWVYGPNGNEKSGQFGKLGKKDYLPGKTKDDMTYMLMKNPKANVYMD  
 IPRCNSEYLNQFMELIKNRTIFSYKPEYVGCIIINNKIHIIVILANVLPDYKISQDRI  
 KIIYC"  
 939. .944  
 /gene="C9-Eg"  
 998. .1003  
 /gene="C9-Eg"  
 353 a 161 c 227 g 266 t  
 BASE COUNT  
 ORIGIN

Query Match 31.8%; Score 89.4; DB 58; Length 1007;  
 Best Local Similarity 59.1%; Pred. No. 2.7e-14;  
 Matches 153; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
 QY 7 tgggtgaaccaggaagacgtattgtaccagtcagagaccacacgaacgaatttaatt 66  
 Db 660 TGGAGGTAAACCCCAAGATATGACATATATGTTAAATGAAATCCAAAGCAATGTTGT 719  
 QY 67 cctcgatgaccaggtgatttagagatttaattgacccctgttagaattgttta 126  
 Db 720 GATGATATTCCTCGTTGTAATTCGAATATTTAAATATATCAATTTATGGAATTAATTA 779

QY 127 gaacagggcattcagttcggacaaatacgaaccccttagttatcttgggtcgaccatgt 186  
 Db 780 AATAGAACCATATTTAGTTATAAATATGAACCCAGTTGGATGATTATAATAATAAAT 839  
 QY 187 gcatgtactctatttgcgaatgtcctgcctgattattgaaatcagcagcagacgaat 246  
 Db 840 ACATGTAATTTGATTAGCTAATGATTATTCCTGATTGCAAAAATTAAGTCAGGACAGAAT 899  
 QY 247 aaaaactgtggaatttttaa 265  
 Db 900 TAAATAATTTATTGTTAA 918  
 RESULT 3  
 NVV5966  
 LOCUS  
 DEFINITION faba bean necrotic yellows virus C9 gene. 13-JAN-1999  
 ACCESSION AJ005966  
 VERSION AJ005966.1 GI:3550532  
 KEYWORDS C9 gene; component 9; putative; rep protein; replication associated protein.  
 SOURCE  
 ORGANISM  
 faba bean necrotic yellows virus.  
 Viruses; ssDNA viruses; Nanovirus.  
 1 (bases 1 to 1004)  
 Katul, L.  
 Direct Submission  
 Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und  
 Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und  
 Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY  
 2 (bases 1 to 1004)  
 Katul, L., Timchenko, T., Gronenborn, B. and Vetten, H.J.  
 Ten distinct circular ssDNA components, four of which encode  
 putative replication-associated proteins, are associated with the  
 faba bean necrotic yellows virus genome  
 J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)  
 99094637  
 Location/Qualifiers  
 1. .1004  
 /organism="faba bean necrotic yellows virus"  
 /virion  
 /isolate="SV292-88"  
 /db\_xref="taxon:59817"  
 73. .918  
 /gene="C9"  
 73. .918  
 /gene="C9"  
 /function="putative replication associated (rep) protein"  
 /product="component 9"  
 /protein\_id="CAA06789.1"  
 /db\_xref="GI:3550533"  
 /db\_xref="SPTREMBL:Q91252"  
 /translation="MSAVNVVFTLNFAGEVPVLSFDERVOYAVWQHRYVNHDIQGV  
 OLKKAKNTVKNIIIGGNPHLEKMGSIIEASAYAOKEESRVAGPWSYGEGLKKGSHK  
 RKIMELIKDPENLEERPKYBRAMANSAMDESKLAEEGGFPYMFYSWQETVLGLLEE  
 EPNDRTIIWVYGPNGNEKSGQFGKLGKKDYLPGKTKDDMTYMLMKNPKANVYMD  
 IPRCNSEYLNQFMELIKNRTIYSKYEPVGCIIINNKIHIIVILANVLPDYKISQDRI  
 KIIYC"

FEATURES  
 Source

gene  
 CDS

BASE COUNT 352 a 159 c 226 g 267 t  
 ORIGIN

Query Match 30.7%; Score 86.2; DB 59; Length 1004;  
 Best Local Similarity 58.3%; Pred. No. 2e-13;  
 Matches 151; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
 QY 7 tgggtgaaccaggaagacgtattgtaccagtcagagaccacacgaacgaatttaatt 66  
 Db 660 TGGAGGTAAACCCCAAGATATGACATATATGTTAAATGAAATCCAAAGCAATGTTGT 719  
 QY 67 cctcgatgaccaggtgatttagagatttaattgacccctgttagaattgttta 126

QY 1 cacatgtgtggaaccagaaagacgtattgtaccagttacatcgcagggcccccaaaacgaaa 60  
Db 672 CACGCCGCGAGGAGAAAGATGACGTGGCTTATAGTTTGTGGAGGACCCCAAGGCA 731  
QY 61 tttaactcctcgatgtaccaggtgtaatttagagattttaaatattgcccctgttagaatg 120  
Db 732 TGTGGTATTGTATATACACGATGATGCAAGAAATATGTAATACAGCTTAAATAGAGAT 791  
QY 121 tgttaagaacagagcattcagttcgcgaacaatacgaaccccttagttatcttgggttcga 180  
Db 792 GCTTAAGGATAGAAATATTATTAGTACAAATATGAGCAATTTACAAATGCAAGTGA 851  
QY 181 ccatgtgcattactcgtatttgcgaatgtccctgcctgattatttgaataatcagcaggga 240  
Db 852 TAATATTCATGTAATTTGTAATGCTAACTTTTACCTGATGTAACAAATAATATCTGAAGA 911  
QY 241 cagaataaaactgtggaatattttaaaagtattgt 274  
Db 912 TAGAATAAAATAATTTATTGTTGAAATTCGCGT 945  
RESULT 5  
AR063453 1091 bp DNA PAT 29-SEP-1999  
LOCUS Sequence 5 from patent US 5846705.  
DEFINITION AR063453  
ACCESSION AR063453  
VERSION AR063453.1 GI:5992761  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1091)  
AUTHORS Wu, R., You, L. and Soong, T.  
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus.  
JOURNAL Patent: US 5846705-A 5 08-DEC-1998;  
FEATURES Location/Qualifiers  
source 1..1091  
BASE COUNT 343 a 224 c 248 g 276 t  
ORIGIN  
Query Match 25.2%; Score 70.8; DB 9; Length 1091;  
Best Local Similarity 56.2%; Pred. No. 3.4e-09;  
Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;  
QY 8 ggtggaaccagaaagacgtattgtaccagttacatcgcagggcccccaaaacgaaatttaac 67  
Db 663 GGTGGAAGACGTCGGATATGATGCACATCATACGATGATGATGATGATGATGATGATGAT 722  
QY 68 ctcatgtaccaggtgtaatttagagattttaaatattgcccctgttagaatgtgtaag 127  
Db 723 ATTGATATCCCGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 782  
QY 128 acagggcattcagttcgcgaacaatacgaaccc---ccttagttatcttgggttcgacat 184  
Db 783 AATAGAGTTTAAATAATACAAATACGACCATGTCATGATGATGATGATGATGATGATGATGAT 842  
QY 185 gtcatgtactcgtatttgcgaatgtccctgcctgattatttgaataatcagcagggaacaga 244  
Db 843 GTCCATGTAATGTTTATGCAAAATGTTGCTGATGATGATGATGATGATGATGATGATGATGAT 902  
QY 245 ataaactgtggaatattttaaaagtattgtgtc 278  
Db 903 ATAAAAATAATTAATTTGTTGAGAAAGGAAATTC 936  
RESULT 6  
AR063452 1096 bp DNA PAT 29-SEP-1999  
LOCUS Sequence 4 from patent US 5846705.  
DEFINITION

Db 720 GATGGATATTCCTGCTTAATTTCTGAATATTAAATACCAATTTATPGGAATTAATAA 779  
QY 127 gaacagggcattcagttcgcgaacaatacgaaccccttagttatcttgggttcgaccatgt 186  
Db 780 AAATAGAACCATATATAGTTTAAATATGACACGATGATGATGATGATGATGATGATGATGAT 839  
QY 187 gcatgtactcgtatttgcgaatgtccctgcctgattatttgaataatcagcagggaacagaat 246  
Db 840 ACATGTAATGTTAGTAAATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 899  
QY 247 aaaactgtggaatatttaa 265  
Db 900 TAAATAAATTTATTGTTAA 918  
RESULT 4  
AB000920 1007 bp DNA circular VRL 30-OCT-1998  
LOCUS Milk vetch dwarf virus genome segment 1 encoding viral  
DEFINITION replication-associated protein, complete sequence.  
ACCESSION AB000920  
VERSION AB000920.1 GI:3808179  
KEYWORDS viral replication-associated protein.  
SOURCE milk vetch dwarf virus (lab\_host:Pisum sativum) DNA.  
ORGANISM milk vetch dwarf virus  
VIRUSES; ssDNA viruses; Nanovirus.  
REFERENCE 1 (bases 1 to 1007)  
AUTHORS Sano, Y.  
TITLE Direct Submission  
JOURNAL Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.  
Yoshitaka Sano, Kyoto Institute of Technology, Department of Applied Biology; Matsugasaki, Sakyo-Ku, Kyoto 606-8585, Japan (Tel:075-724-7764; Fax:075-724-7764)  
2 (sites)  
Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.  
AUTHORS Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome  
JOURNAL J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)  
MEDLINE 99094638  
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798651.  
Sequence updated (30-Sep-1997).  
FEATURES Location/Qualifiers  
source 1..1007  
/organism="milk vetch dwarf virus"  
/viral  
/db\_xref="taxon:67585"  
/chromosome="segment 1"  
/lab\_host="Pisum sativum"  
1..33  
stem\_loop 51..57  
TATA\_signal 91..936  
CDS  
/function="helicase"  
/note="putative"  
/codon\_start=1  
/product="viral replication-associated protein"  
/protein\_id="BAA33980.1"  
/db\_xref="GI:3798652"  
/translation="MPTLQGTFCWFTLNFSGDAPSLSENERVQACWHERVSHDLQ  
GYOMKRSITKMKKLELGAHLEYSKVPPEASDYAMKEETRVAGPTWYGLKKKGS  
NKKRLDRYKENPEDEWLEDPARAKRKAIDKEKFAEFKVEDDEQWKKILEKEIE  
KIAPSRILWYGPQGGEGTSKAKELITGWFYTRGGRKDDVAYSYVEDPRHYVFD  
IPRDMQBYCNYSLIEMLKDRILISNKEPITNQVYNIHIVNMLPVDVTKISEDR  
KIIVC"  
polyA\_signal 915..920  
BASE COUNT 348 a 177 c 230 g 252 t  
ORIGIN  
Query Match 26.3%; Score 74; DB 58; Length 1007;  
Best Local Similarity 54.4%; Pred. No. 4.5e-10;  
Matches 149; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

ACCESSION AR063452  
 VERSION AR063452.1 GI:5992760  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1096)  
 AUTHORS Wu, R., You, L., and Soong, T.  
 TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus  
 JOURNAL Patent: US 5846705-A 4 08-DEC-1998;  
 FEATURES Location/Qualifiers  
 source 1..1096  
 BASE COUNT 347 a 231 c 244 g 274 t  
 ORIGIN  
 Query Match 25.2%; Score 70.8; DB 9; Length 1096;  
 Best Local Similarity 56.2%; Pred. No. 3.4e-09;  
 Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;  
 Qy 8 ggtggaaccagaagcgtatgtaccagtagcatatgacatgagagaccacaaacgaatttaac 67  
 Db 668 GGTGGAAGACGTCGGATATGATGCACATCATACACGATGATCCTGATTAATCATTTGGATT 727  
 Qy 68 ctgatgtaccagtgtaattagatgtatttaattatgcctgttgaagaatgtttaag 127  
 Db 728 ATTGATATCCCAAGTCATTCAGATTATCTCAATTATGCGCTTATAGAACAAATTAAG 787  
 Qy 128 aacaggcattcagttcgagacaaatcgaacc---ccttagttattcttgggttcgacct 184  
 Db 788 AATAGAGTTTATAATAATACAAATACGAACCATGTGTGATTAGAAAAGATGGACAAAT 847  
 Qy 185 gtgcattgactctgtattgccaatgctgcctgattatttgaataatcagcaggagacaga 244  
 Db 848 GTCCATGTAATTTGTTATGCAATGTGTCCTGATTAATTTGTAATAATTTTCAGAAGATAGA 907  
 Qy 245 ataaactgtggaatttaagtagtgcac 278  
 Db 908 ATAAAAATAATTAATTTGTGAGAAAGGAAACTTC 941  
 RESULT 7  
 LOCUS BYTV2 1096 bp. DNA VRL 30-OCT-1994  
 DEFINITION Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes, complete cds's.  
 ACCESSION L32167  
 VERSION L32167.1 GI:520791  
 KEYWORDS stem loop.  
 SOURCE Banana bunchy top virus DNA.  
 ORGANISM Banana bunchy top virus  
 Viruses; ssDNA viruses; Circoviridae.  
 REFERENCE 1 (bases 1 to 1096)  
 AUTHORS Wu, R.-Y., You, L.-R., and Soong, T.-S.  
 TITLE Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus  
 JOURNAL Unpublished (1994)  
 FEATURES Location/Qualifiers  
 source 1..1096  
 /organism="Banana bunchy top virus"  
 /db\_xref="taxon:12585"  
 TATA\_signal 1..7  
 stem\_loop 71..38  
 gene 71..928  
 /gene="V2"  
 CDS 71..928  
 /gene="V2"  
 /codon\_start=1  
 /protein\_id="AAA51426.1"  
 /db\_xref="GI:520792"

/translation="MSEFKWCFTLYSSAAREDFLALLKEEDVHYSVVGDEVAPATG  
 OKHOGVLSLKKSIIRLGGILKKYKSAHWEIAKGSDEQNRRCYSKETLVLEIGTPVP  
 GSKRLLDRFESPEELKMDPSKYRCLAVESKDARINSEWHEIKWQNKLIQH  
 IEGVDDRSIIWYVPGNGEGKSTFARYLSLKPWGVI NGKTSMDMHIITMDPNHWH  
 IEDPRSHSDIYNGVIEQIKNRVLINTRYEPCVIRKQGNVHVIMANVLDPYCKIS  
 EDRIKIIINC"  
 85..309  
 /gene="V1"  
 85..309  
 /gene="V1"  
 /codon\_start=1  
 /protein\_id="AAA51427.1"  
 /db\_xref="GI:520793"  
 /translation="MVLHSELFLRGARLRSRSGGCSLLCRRRRSRSGHRPEAPP  
 GSIPEKINSPPRIEKEVWLPCLGDCERK"  
 complement(544..699)  
 /gene="C2"  
 complement(544..699)  
 /gene="C2"  
 /codon\_start=1  
 /protein\_id="AAA51428.1"  
 /db\_xref="GI:520794"  
 /translation="MMCIIISDVFPPLIYPHGFNDRLAKVDFPSPPLGVPYQMILRS  
 SGTPTSMC"  
 complement(883..1059)  
 /gene="C1"  
 complement(883..1059)  
 /gene="C1"  
 /codon\_start=1  
 /protein\_id="AAA51429.1"  
 /db\_xref="GI:520795"  
 /translation="MGLRPNKRFINYNDFLMHLMIACSASYRVGSTCTFLIACGGSFLS  
 QQLIIFILSSEILQ"  
 BASE COUNT 347 a 231 c 244 g 274 t  
 ORIGIN  
 Query Match 25.2%; Score 70.8; DB 58; Length 1096;  
 Best Local Similarity 56.2%; Pred. No. 3.4e-09;  
 Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;  
 Qy 8 ggtggaaccagaagcgtatgtaccagtagcatatgacatgagagaccacaaacgaatttaac 67  
 Db 668 GGTGGAAGACGTCGGATATGATGCACATCATACGATGATCCTGATTAATCATTTGGATT 727  
 Qy 68 ctgatgtaccagtgtaattagatgtatttaattatgcctgttgaagaatgtttaag 127  
 Db 728 ATTGATATCCCAAGTCATTCAGATTATCTCAATTATGCGCTTATAGAACAAATTAAG 787  
 Qy 128 aacaggcattcagttcgagacaaatcgaacc---ccttagttattcttgggttcgacct 184  
 Db 788 AATAGAGTTTATAATAATACAAATACGAACCATGTGTGATTAGAAAAGATGGACAAAT 847  
 Qy 185 gtgcattgactctgtattgccaatgctgcctgattatttgaataatcagcaggagacaga 244  
 Db 848 GTCCATGTAATTTGTTATGCAATGTGTCCTGATTAATTTGTAATAATTTTCAGAAGATAGA 907  
 Qy 245 ataaactgtggaatttaagtagtgcac 278  
 Db 908 ATAAAAATAATTAATTTGTGAGAAAGGAAACTTC 941  
 RESULT 8  
 LOCUS AR063451 1106 bp. DNA PAT 29-SEP-1999  
 DEFINITION Sequence 3 from patent US 5846705.  
 ACCESSION AR063451  
 VERSION AR063451.1 GI:5992759  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1106)

[illegible]

|            |   |
|------------|---|
| RESULT     | 10  |
| AF216222   |   |
| LOCUS      | 1095 bp DNA 08-MAR-2001   |
| DEFINITION | Banana bunchy top virus satellite S2 replication initiation protein<br>(ORF V1) gene, complete cds. |
| ACCESSION  | AF216222  |
| VERSION    | AF216222.1 GI:12004327  |
| KEYWORDS   | .   |
| SOURCE     | Banana bunchy top virus.  |
| ORGANISM   | Banana bunchy top virus   |
| REFERENCE  | Viruses; ssDNA viruses; Circoviridae.<br>1 (bases 1 to 1095)  |
| AUTHORS    | Horsefield,C.L., Karan,M., Harding,R.M. and Dale,J.L.   |
| TITLE      | Additional rep-encoding DNAs associated with banana bunchy top<br>virus                             |
| JOURNAL    | Arch. Virol. 146 (1), 71-86 (2001)  |
| MEDLINE    | 21163975  |
| REFERENCE  | 2 (bases 1 to 1095)   |







---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:47 ; Search time 5701.85 Seconds  
(without alignments)  
1033.561 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_611\_991

Perfect score: 381

Sequence: 1 cagcgagcatcgacacatc.....ttaaagtatgtgtcatctaa 381

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_bal.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_bal.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_hgl.\*  
61: gb\_hgl2.\*  
62: gb\_hgl3.\*  
63: gb\_hgl4.\*  
64: gb\_hgl5.\*  
65: gb\_hgl6.\*  
66: gb\_hgl7.\*  
67: gb\_hgl8.\*  
68: gb\_hgl9.\*  
69: gb\_hgl10.\*  
70: gb\_hgl11.\*  
71: gb\_hgl12.\*  
72: gb\_hgl13.\*  
73: gb\_hgl14.\*  
74: gb\_hgl15.\*  
75: gb\_hgl16.\*  
76: gb\_hgl17.\*  
77: gb\_hgl18.\*  
78: gb\_hgl19.\*  
79: gb\_hgl20.\*  
80: gb\_hgl21.\*  
81: gb\_hgl22.\*  
82: gb\_hgl23.\*  
83: gb\_hgl24.\*  
84: gb\_hgl25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rol2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 381   | 100.0       | 1291   | 58    | CFDCG              |
| 2          | 115.8 | 30.4        | 1007   | 58    | FBE132187          |
| 3          | 114.2 | 30.0        | 1004   | 59    | NYV5966            |
| 4          | 112.4 | 29.5        | 1096   | 9     | AR063452           |
| 5          | 112.4 | 29.5        | 1096   | 58    | BYTV2              |
| 6          | 110.8 | 29.1        | 1091   | 9     | AR063453           |
| 7          | 110.6 | 29.0        | 1095   | 58    | AF216222           |
| 8          | 96    | 25.2        | 1111   | 58    | BBU12586           |
|            |       |             |        |       | M29963 Coconut fol |
|            |       |             |        |       | AJ132187 faba bean |
|            |       |             |        |       | AJ005966 faba bean |
|            |       |             |        |       | AR063452 Sequence  |
|            |       |             |        |       | L32167 Banana bunc |
|            |       |             |        |       | AR063453 Sequence  |
|            |       |             |        |       | AF216222 Banana bu |
|            |       |             |        |       | U12586 Banana bunc |





— — — — —

```
/codon_start=1
```

```

/protein_id="AA51429.1"
/db_xref="GI:520795"
/translation="MGLRPKRFNYNDFLHMLIACSAASYRVGSTCTLIACGGFSL
QQLIFLSLEILQ"
347 a      231 c      244 g      274 t
BASE COUNT
ORIGIN

```

Query Match 29.5%; Score 112.4; DB 58; Length 1096;  
Best Local Similarity 57.9%;  
pred. No. 7.2e+22;  
Matches 219; Conservative 0; Mismatches 156; Indels 3; Gaps 1;

|    |     |  |     |
|----|-----|--|-----|
| Qy | 4   | cggacgagtcgcacaatctctcgatctgcgcagacgcgagggagcgcgaagtcccggt       | 63  |
| Db | 564 | CTGATGATCGAAGTATCATCTGGGTATACGGTCCCAACGGAGGCCAAGGAAGTCAACCT      | 623 |
| Qy | 64  | ttgccaaatatctcgcgactcaagcccgactggtttcacacatggttggaaccagaagaag    | 123 |
| Db | 624 | TGCGAAGATATCTATCATTAAGCCCGATGGGATATATCAACGGTGGAAAGACGTCGG        | 683 |
| Qy | 124 | acgtattgtaccagtcacatcgagaccccaaacgaaatttaactcctcgatgtagtaccagggt | 183 |
| Db | 684 | ATATGATGCCATCATTAACGATGGATCCTGATAATCATGATATTGATATCCCCAGAA        | 743 |
| Qy | 184 | gtaatttagagtgatttaaatattgcccctgttagaattgtttaagaacaggcgattcaggt   | 243 |
| Db | 744 | GTCATTCAGATTATCTGAATTATGCGGTATAGNACAAATTAGAATAGAGTTTTAATAA       | 803 |
| Qy | 244 | cggacaaatacgaacc--ccctagtattcttgggttcgaccattgtgcattgtactcgat     | 300 |
| Db | 804 | ATACAAATACGAACCATGTGTGATTAGAAAGATGACAAAATGTCCTATGTAATTGTTA       | 863 |
| Qy | 301 | ttgccaatgtctcctcctgattattgaaatacagcagggcacagaataaactgtggaata     | 360 |
| Db | 864 | TGGCAATGTCTGCCCTGATTATTGTAATAATTCAGAGATAGATAAAAAATAATTAAT        | 923 |
| Qy | 361 | tttaaagtatgtgtcatc   | 378 |
| Db | 924 | GTTGAGAAAGGAAACTTC   | 941 |

|            |   |
|------------|---|
| RESULT     | 6   |
| LOCUS      | AR063453  |
| DEFINITION | Sequence 5 from patent US 5846705.  |
| ACCESSION  | AR063453  |
| VERSION    | AR063453.1 GI:5992761   |
| KEYWORDS   | Unknown.  |
| SOURCE     | Unknown.  |
| ORGANISM   | Unclassified.   |
| REFERENCE  | 1 (bases 1 to 1091)   |
| AUTHORS    | Wu, R., You, L. and Soong, T.   |
| TITLE      | Nucleotide sequence of two circular SSDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus |
| JOURNAL    | Patent: US 5846705-A 5 08-DEC-1998;   |
| FEATURES   | Location/Qualifiers<br>1..1091<br>/organism="unknown"   |
| BASE COUNT | 343 a 224 c 248 g 276 t   |
| ORIGIN     |   |

Query Match 29.1%; Score 110.8; DB 9; Length 1091;  
Best Local Similarity 57.7%; Pred. No. 2.1e-21;  
Matches 218; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

Qy 4 cggacgatcgacaatcctctggtatatcgacagacgaggagacgggaagtccctgt 63  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 559 CTGATGATCGAAGTATCATCTGGGTATACGGTCTCTCCGGAGCGCAAGGAAGTCAACT 618  
Qy 64 tlgccaataatctcggactcaagcccgactggttctacacatgttggtggaccagaaagg 123

|            |   |   |             |
|------------|---|---|-------------|
| Db         | 619   | TCCGACAGATATCATCATTTAAACCTGGATGGGGATATATCAACGGTGGAAAGACGTCGG    | 678         |
| QY         | 124   | acgtattgtaccagtgacatcgagagaccgaacgaataatttaacctcgtggtaccaggt    | 183         |
| Db         | 679   | ATATGATGCCACATCATAAACGATGGATCCTGATATAATCATTCGATTATGATATCCCCAGAA | 738         |
| QY         | 184   | gtaattagaagtattaaattatgccctgttaagaatggttaagaacacagggcattcagtt   | 243         |
| Db         | 739   | GTCATTCAAGATTATCTGAATTATGGCGTTATAGACAAATTAAGAATAGAGTTTTAATAA    | 798         |
| QY         | 244   | cggacaataacgaacc---ccttagttatcttgggtctgaccatgtgcctgtaclogtat    | 300         |
| Db         | 799   | ATACAAAATACGAACCATGTGTGATTAGAAAAGATGGCAAAATGTCCAATGTAATGTTA     | 858         |
| QY         | 301   | ttagcaatgtcctgctgattatttgaataatcagcaggagacagaataaaactgtggaata   | 360         |
| Db         | 859   | TGCAAAATGTGTGCTGATTATTTGTAATTCAGAGATAGATAAAAAATAATTAATT         | 918         |
| QY         | 361   | tttaaagtagtgtgtcatc   | 378         |
| Db         | 919   | GTTGAGAAAGGAACATTC  | 936         |
| RESULT     | 7   |   |             |
| LOCUS      | AF216222  | 1095 bp DNA VRL   | 08-MAR-2001 |
| DEFINITION | Banana bunchy top virus satellite S2 replication initiation protein (ORF V1) gene, complete cds.                                  |   |             |
| ACCESSION  | AF216222  |   |             |
| KEYWORDS   | AF216222.1  | GI:12004327   |             |
| SOURCE     | Banana bunchy top virus.  |   |             |
| ORGANISM   | Banana bunchy top virus   |   |             |
| REFERENCE  | Viruses; ssDNA viruses; Circoviridae.   |   |             |
| AUTHORS    | 1 (bases 1 to 1095)   |   |             |
| TITLE      | Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.   |   |             |
| JOURNAL    | Additional rep-encoding DNAs associated with banana bunchy top virus  |   |             |
| MEDLINE    | Arch. Virol. 146 (1), 71-86 (2001)  |   |             |
| REFERENCE  | 21163975  |   |             |
| AUTHORS    | 2 (bases 1 to 1095)   |   |             |
| TITLE      | Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.   |   |             |
| JOURNAL    | Direct Submission   |   |             |
| FEATURES   | Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia |   |             |
| source     | Location/Qualifiers   |   |             |
| mRNA       | 1..1095   |   |             |
| gene       | /organism="Banana bunchy top virus"   |   |             |
| CDS        | /isolate="Taiwan"   |   |             |
|            | /db_xref="taxon:12585"  |   |             |
|            | /note="satellite S2"  |   |             |
|            | 64..921   |   |             |
|            | /gene="ORF V1"  |   |             |
|            | 64..921   |   |             |
|            | /gene="ORF V1"  |   |             |
|            | 64..921   |   |             |
|            | /gene="ORF V1"  |   |             |
|            | /codon_start=1  |   |             |
|            | /product="replication initiation protein"   |   |             |
|            | /protein_id="AAG44004.1"  |   |             |
|            | /db_xref="GI:12004328"  |   |             |
|            | /translation="MSSPFWKCTLNYSSAAAREDFELALLKEELNAYVVGDEVPASTGRKHQGLYSLKSKIKLGLLKKYSSKAHWERAGDEQNRGYCSKETILVEIGTPVP                   |   |             |
|            | GSKKKLLRFRFEPSELKMEQPSKYRCLAVESLNNARKNSEWHLEKRWONKLIQH  |   |             |
|            | IEGVDDRSIIWYGPNGEGEKSTFARYLSLKPWGVIYNGKTSMDMHIITMDPNHW  |   |             |
|            | IDILPRSDYLNIGYVIEQIKNRVLINTKYPCVIRKQCNVHVIMANVLPDYCKIS  |   |             |
|            | EDRIKIING"  |   |             |
| BASE COUNT | 347 a   | 219 c   | 277 t       |
| ORIGIN     |   |   |             |

[illegible]

| Query Match           | 24.9%        | Score 95  | DB 58          | Length 1002 |        |
|-----------------------|--------------|---|----------------|-------------|--------|
| Best Local Similarity | 55.6%        | Pred. No. 7.6e-17   |                |             |        |
| Matches 204           | Conservative | 0   | Mismatches 160 | Indels 3    | Gaps 1 |
| QY                    | 2            | agcggacgacgcacaaatcctctgdatatgcgacgacgagcggagacggaagtcogt       | 61             |             |        |
| Dbb                   | 639          | ACCAGATGATCGTACTATNCATCTGGTTTATGGACCAGATGGAGGAGAAGAAAATCGAT     | 698            |             |        |
| QY                    | 62           | gtttgcgaatatctcggactcaagccgactggtctacacatgtggtggaaccagaaa       | 121            |             |        |
| Dbb                   | 699          | TTTCGCGAAGAATTT--AATCACTATGGATGGTTTTTACACTGCCGGAGGAAGACGCCA     | 755            |             |        |
| QY                    | 122          | ggacgtattgtaccagtcacgcgaggaccaccaaaacgaatttaactctcctgatgtaccacg | 181            |             |        |
| Dbb                   | 756          | GGACATCTGTATGTATGTCTCAAGATCCAGAGAGAAATATTGCATTTGTATGTATCCCG     | 815            |             |        |
| QY                    | 182          | gtgtaatttagagctattaaattatgcccctgttagaatgtgtaagacagggcattcag     | 241            |             |        |
| Dbb                   | 816          | ATGTTTCATCCGAAATGATGAACTATCAAGCGATGGAGATGATGAAGAATAAGATGCTTTGC  | 875            |             |        |
| QY                    | 242          | ttcggacaaatcacgaaccccttagtctatctgggttcgaccatgtcgatgaactcgtatt   | 301            |             |        |
| Dbb                   | 876          | AAGTACGAAATATAGATCTGTAGATCTTTGTGTATATAAAATGTTCATTAGTTGTTTT      | 935            |             |        |
| QY                    | 302          | tgccaatgtcctgcctgattatttggaaatcacgacgggcacagaataaaactgtggaatat  | 361            |             |        |
| Dbb                   | 936          | TGCCAACGTGGCATATGACCCCAAAAATAAGTGGAGATAGGATTGTAATTATCATCAATG    | 995            |             |        |

```

FEATURES             source
Location/Qualifiers
  1. .1127
    /organism="Banana bunchy top virus"
    /db_xref="taxon:12585"
    /clone="S15 and 4-2"
  1. .8
    15. .49
    79. .933
    /note="ORF V1"
    /codon_start=1
    /protein_id="AAA61877.1"
    /db_xref="GI:642394"
    /translation="MSSFQWCTFLNYSAAEREDLALLKEELNVAVWGDEVAPSSG
    QKHGGVLSKKSIKIGGLKKYSSRAHWERAGSDENAKYCKETLILLEGPASQ
    GSNRKLSEWSPSPRMRIQOEPIYHRYTSYKLLKKEEFVHPCLDRWQIQLTFA
    IDEPDDRTLFWYQNGEGSKTYAKSLMKKDWFYTRGGKKENILFSYVDEGSEKHI
    VFDPICNQDLYNDVIEALKDORVISTKYKPKIKVLNNTHVIVMANFMPEFKISE
    DRKILYYC"
  complement(497. .643)
    /note="ORF C1"
    /codon_start=1
    /protein_id="AAA61878.1"
    /db_xref="GI:642395"
    /translation="MTSHMLISPHYHSDHRRPRSFEDHLVPRQLPPSIESAMVYRGKDE
    RIP"
353 a      222 c      260 g      292 t
BASE COUNT
          251C1N

```

| Query Match           | 24.8%;   | Score 94.4;  | DB 58;    | Length 1127; |
|-----------------------|--|--|-----------|--------------|
| Best Local Similarity | 57.3%;   | Pred. No. 1.le-16;   |           |              |
| Matches 211;          | Conservative 0;  | Mismatches 151;  | Indels 6; | Gaps 2:      |
| 2                     | agcgagcagatcgacaaatcctctggaatgcgagacgagacgagagacggagagtcctcg | 61   |           |              |
| db                    | 570  | ACCAGATGATCGAACGATCTCTGGGTCTATGGTCCGAATGCTAATGAGCGGAATCAAC     | 629       |              |
| 2y                    | 62   | gtttgcacaaatctctcgactcaagccgagactggtctacacatgtagtgagaccgagaaa  | 121       |              |
| db                    | 630  | ATATGCGAAGCTCATTT---AATGAGAAAGAGACTGGTCTTACACCAGAGTGGGAGGAAGGA | 686       |              |
| 2y                    | 122  | ggacgtattgtaccagatcagtaggaccccaaacgaa---atttaactctcgatgacc     | 178       |              |
| db                    | 687  | GAACATACTGTTCTCCTACGTCGACGAGGATCTGAGAAAGCATTTGTAATTGATATTCC    | 746       |              |

1

1

1





Thu Oct 25 13:08:19 2001

```

MEDLINE      21102983
PUBMED       11161286
REFERENCE    2 (bases 1 to 1109)
AUTHORS      Horser, C.L., Karan, M., Harding, R.M. and Dale, J.L.
TITLE        Additional rep-encoding DNAs associated with banana bunchy top
            virus
JOURNAL      Arch. Virol. 146 (1), 71-86 (2001)
REFERENCE    3 (bases 1 to 1109)
AUTHORS      Horser, C.L., Karan, M., Harding, R.M. and Dale, J.L.
TITLE        Direct Submission
JOURNAL      Submitted (14-DEC-1999) School of Life Sciences, Queensland
            University of Technology, GPO Box 2434, Brisbane, QLD 4001,
            Australia
FEATURES     Location/Qualifiers
            1..1109
            /organism="Banana bunchy top virus"
            /isolate="Taiwan"
            /db_xref="taxon:12585"
            /note="satellite S1"
            3..33
            /note="putative sequence"
            63..917
            /gene="ORF V1"
            63..917
            /gene="ORF V1"
            63..917
            /gene="ORF V1"
            /codon_start=1
            /product="replication initiation protein"
            /protein_id="AAG44003.1"
            /db_xref="GI:12004326"
            /translation="MSSFKWCFILNYSRAEREDFLALKEELNVAVGVDEVPSSG
            QKHLQGYLSKSLKGLKYLSSRAHWRARGSDNAKYSKEILLLELGPASQ
            GSNRKLSEVSRPEMRIEQPEIYHRYTSVAKLKKFEEFVHPCDLRPOQLTEA
            IDEPDDRSIIWVYCPNGNEGKSTYAKSLMKDWFYRGCKENILFVYDGESEKHI
            VFDIPRCNODLYNDVIEALKDRVIESTKYKPIKLVELINIHVIMANFEPCKISE
            DRKILYC"
BASE COUNT   346 a 213 c 257 g 293 t
ORIGIN
Query Match      23.9%; Score 91.2; DB 58; Length 1109;
Best Local Similarity 56.8%; Pred. No. 9.5e-16;
Matches 209; Conservative 0; Mismatches 153; Indels 6; Gaps 2;

Qy 2 agcggacatcgacaaatcctctgatatcgaggacgacgagagacggaatccgt 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 ACCCGATGATCGAAGCATCATCTGGTCTATGGTCCGAATGTAATGAGGGGAATCAAC 613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 62 gttgcaaatatctcggaactcaagcccgactggttctacacatgtgtggaaccagaaa 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 ATATGCGAAGTCAATT---AATGAGAGAGACGTGGTTCTACACACAGGTGGGAAGGA 670
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 122 ggacgtattgtaccagatcatcgagagaccacaaacgaa---attaatcctcgatgtacc 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 GAACATACTGTTCTTCTACGTGACGACGAAGGATCTCAGAACGATATTGTATTGTATTC 730
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 179 caggtgtaattagagatttaaatattaccctgttagaatgtgttaagaacagggcatt 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 731 TCGTGTGAATCAGGATTATTAAATATTGATTGTTATAGAGGATTAAAGATAGGTTGAT 790
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 239 cagttcggaacaaatacgaaccccttatcttctggttgcaccatgtgcatgtactcgt 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 791 AGACAGTACTAAATATAAACCTTATAGTGTAGTTGAATTGATTATATACATGTAATTGT 850
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 299 atttgcaatgtcgtcgtcgtatttgaataatcagcagggagacagataaaactgtgaa 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 851 CATGGCTAAATTCATGCCAGATTCTGTAAATCTCCGAAGATAGATAAAGATTATTTA 910
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 359 tatttaaa 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 911 TGTGTTAA 918

```

```

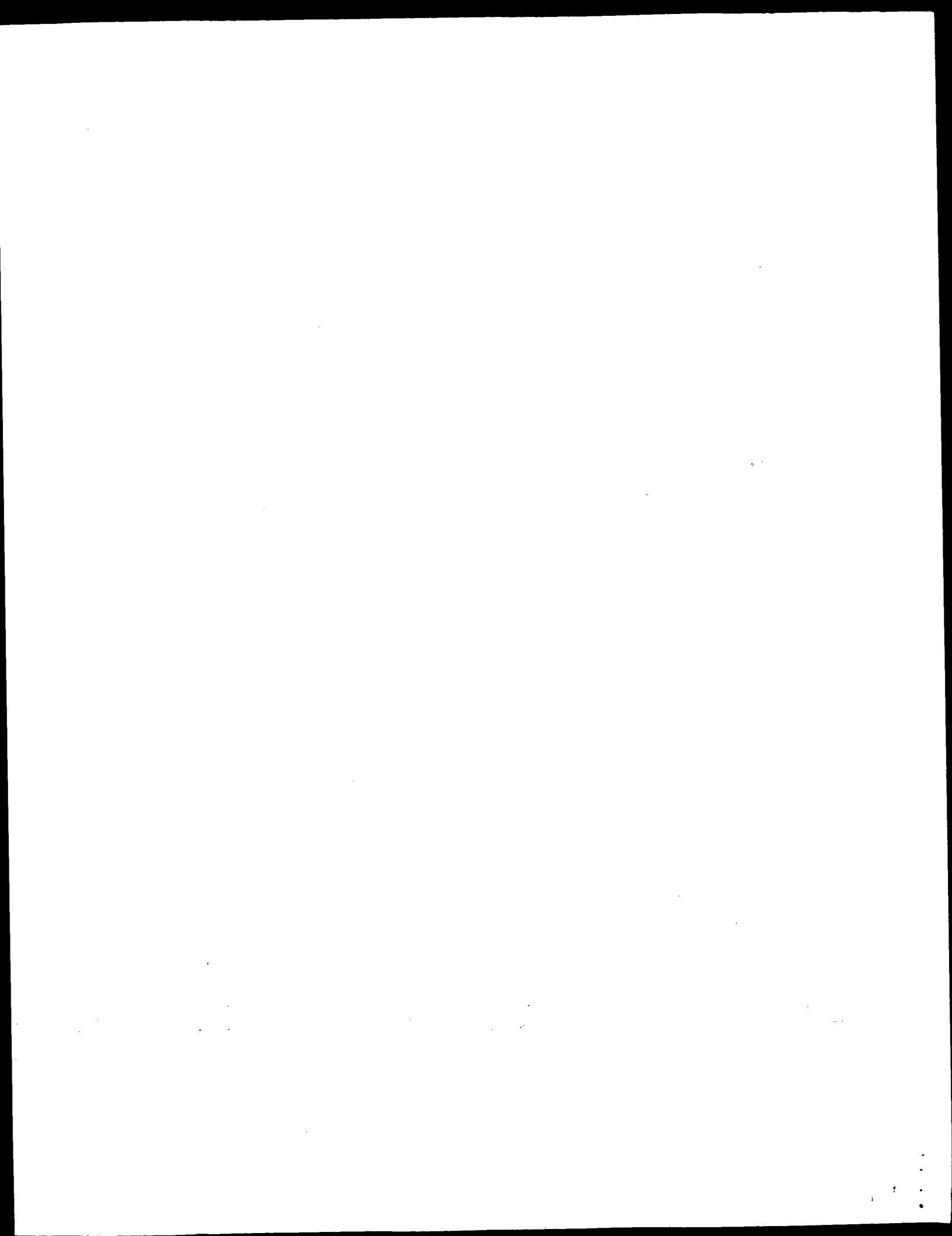
RESULT 15
FBI32185 1015 bp DNA circular VRL 16-SEP-2000
LOCUS faba bean necrotic yellows virus C7-Eg gene, isolate Egyptian
DEFINITION EV1-93.
ACCESSION AJ132185 GI:4995167
VERSION AJ132185.1
KEYWORDS C7-Eg gene; rep protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus
VIRUSES: ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1015)
AUTHORS Katul, L., Timchenko, T., Gronenborn, B. and Vetten, H.J.
TITLE ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
MEDLINE 99094637
REFERENCE 2 (bases 1 to 1015)
AUTHORS Timchenko, T., de Kouchkovsky, F., Katul, L., David, C., Vetten, H.J. and Gronenborn, B.
TITLE A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
JOURNAL J. Virol. 73 (12), 10173-10182 (1999)
MEDLINE 20027244
REFERENCE 3 (bases 1 to 1015)
AUTHORS Katul, L.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Land-Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. Land-u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
FEATURES Location/Qualifiers
            1..1015
            /organism="faba bean necrotic yellows virus"
            /viroion
            /isolate="Egyptian EV1-93"
            /db_xref="taxon:59817"
            /lab_host="Vicia faba"
            1..33
            /gene="C7-Eg"
            60..1014
            /gene="C7-Eg"
            60..911
            /gene="C7-Eg"
            /function="replication-associated protein"
            /note="33.05 kDa"
            /codon_start=1
            /evidence="experimental"
            /product="Rep protein"
            /protein_id="CAB44025.1"
            /db_xref="GI:4995168"
            /db_xref="SPTREMBL:O9WIK0"
            /translation="MPSIRATHWCFTLNFSGSIPEINWADVOYISWIERVGHDLHQ
            GYIOMNKHVTLKKMKLELPGALHLEMAKAPKAIETCOKESAIAGPWEYGTWISSGSH
            KRKLMEFEDGPEEMKLEDPGLYRCLSRVQMKKIRSCSTWDFLRPQDELKLTIEQ
            EPDDRTILWYVYGHGSGKSAFAKYLTKEGWYTAGGKATDMLYSYSLDPTCHVCI
            IPRCTREIINISVIEQIKNRVINTKYECTIRDDGHNVHIVFCNLFDPDVTRISED
            RIKIINC"
            979..984
            /gene="C7-Eg"
            1009..1014
            /gene="C7-Eg"
            346 a 168 c 228 g 273 t
BASE COUNT
ORIGIN
Query Match      23.6%; Score 89.8; DB 58; Length 1015;
Best Local Similarity 56.3%; Pred. No. 2.4e-15;
Matches 210; Conservative 0; Mismatches 157; Indels 6; Gaps 2;

Qy 2 agcggacgacgcacaaatcctctgatatcgaggacgacgagagacggaatccgt 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 542 ACCAGATGACCGTACAAATTATTGGGTGTATGGCCACATGGAGGAGAGGTAAAAAGCGC 601  
QY 62 gtttgccaaatatctcggaactcaagcccgaciggttctacacatgtgtggaaccagaaaa 121  
Db 602 GTTTCGGAATACITTAACCTTGAAGGAGGTGGTGTATACAGCTGGAGGGAAAGCGAC 661  
QY 122 ggaacgtattgtaccagtaacatcgaggagcccaaaacgaaatttaactcctcgatgtacccag 181  
Db 662 AGATATGCTTTACAGCTATTCAITGGACCCACATGCCAGTATGCATTGATATTTCCAG 721  
QY 182 gtgt---aatttagagatttaaatatgcccgttagaatgtgtaagaacagggcatt 238  
Db 722 ATGTACAGAGAGAATAATTAATTAATTTCTGTAATTGAACAGATTAAAAATAGAGTTAT 781  
QY 239 cagttcgggacaaatacgaaccccttagttatcttgggttcg---accatgtcatgtact 295  
Db 782 TATTAATACAAAATATGAACCTTGACTATTAGAGACGATGGTCATAATGTTTCATGTAT 841  
QY 296 cgtatttgccaaatgctcgtcgtattatttgaataatcagcaggggacagaataaaaactgtg 355  
Db 842 TGTATTTTGTAAATTTTACCTGTGTAACAGAAATCAGTGAAGATAGAGATAAATAAAT 901  
QY 356 gaatattiaaagt 368  
Db 902 AAATGTTGAAAT 914

Search completed: October 24, 2001, 11:42:48  
Job time: 6353 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:44 ; Search time 5701.85 seconds  
(without alignments)  
1581.539 Million cell updates/sec

Title: US-09-462-955-1-copy\_409\_991

Perfect score: 583

Sequence: 1 actgctctgagtgcaaaag.....ttaaagtatgtcatctaa 583

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_ba3:\*
- 4: gb\_in1:\*
- 5: gb\_in2:\*
- 6: gb\_in3:\*
- 7: gb\_om:\*
- 8: gb\_ov:\*
- 9: gb\_pat1:\*
- 10: gb\_pat2:\*
- 11: gb\_ph:\*
- 12: gb\_pl1:\*
- 13: gb\_pl2:\*
- 14: gb\_pl3:\*
- 15: gb\_pl4:\*
- 16: gb\_ba1:\*
- 17: gb\_ba2:\*
- 18: gb\_fun:\*
- 19: gb\_htgo\_hum:\*
- 20: gb\_htgo\_inv:\*
- 21: gb\_htgo\_rod:\*
- 22: gb\_htg\_hum1:\*
- 23: gb\_htg\_hum2:\*
- 24: gb\_htg\_hum3:\*
- 25: gb\_htg\_hum4:\*
- 26: gb\_htg\_hum5:\*
- 27: gb\_htg\_hum6:\*
- 28: gb\_htg\_hum7:\*
- 29: gb\_htg\_hum8:\*
- 30: gb\_htg\_inv1:\*
- 31: gb\_htg\_inv2:\*
- 32: gb\_htg\_other:\*
- 33: gb\_htg\_rod:\*
- 34: gb\_hum1:\*
- 35: gb\_hum2:\*
- 36: gb\_hum3:\*
- 37: gb\_hum4:\*
- 38: gb\_hum5:\*
- 39: gb\_hum6:\*
- 40: gb\_hum7:\*
- 41: gb\_in:\*
- 42: gb\_om:\*
- 43: gb\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_v1:\*

59: gb\_v12:\*

60: gb\_htg1:\*

61: gb\_htg2:\*

62: gb\_htg3:\*

63: gb\_htg4:\*

64: gb\_htg5:\*

65: gb\_htg6:\*

66: gb\_htg7:\*

67: gb\_htg8:\*

68: gb\_htg9:\*

69: gb\_htg10:\*

70: gb\_htg11:\*

71: gb\_htg12:\*

72: gb\_htg13:\*

73: gb\_htg14:\*

74: gb\_htg15:\*

75: gb\_htg16:\*

76: gb\_htg17:\*

77: gb\_htg18:\*

78: gb\_htg19:\*

79: gb\_htg20:\*

80: gb\_htg21:\*

81: gb\_htg22:\*

82: gb\_htg23:\*

83: gb\_htg24:\*

84: gb\_htg25:\*

85: gb\_pr1:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_ro1:\*

95: gb\_ro2:\*

96: gb\_in4:\*

97: gb\_pr10:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description        |
|------------|-------|---------|--------------|-------|--------------------|
| 1          | 583   | 100.0   | 1291         | 58    | CFDCG              |
| 2          | 132   | 22.6    | 1004         | 59    | M29963 Coconut fol |
| 3          | 132   | 22.6    | 1007         | 58    | AJ005966 faba bean |
| 4          | 116.4 | 20.0    | 1096         | 9     | AJ132187 faba bean |
| 5          | 116.4 | 20.0    | 1096         | 58    | AR063452 Sequence  |
| 6          | 114.8 | 19.7    | 1091         | 9     | BYTV2              |
| 7          | 114.4 | 19.6    | 1095         | 58    | L32167 Banana bunc |
| 8          | 108.6 | 18.6    | 1111         | 58    | AR063453 Sequence  |
|            |       |         |              |       | AF216222 Banana bu |
|            |       |         |              |       | U12586 Banana bunc |

## ALIGNMENTS

```

RESULT# 1
CFDCG      1291 bp ss-DNA      circular      VRL      02-AUG-1993
Coconut foliar decay virus, complete genome.
M29963
M29963.1   GI:323306
circular; complete genome.
Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
Coconut foliar decay virus
Viruses; ssDNA viruses; Nanovirus.
1 (bases 1 to 1291)
Rhode,W., Randles,J.W., Langridge,P. and Hanold,D.
Nucleotide sequence of a circular single-stranded DNA associated
with coconut foliar decay virus
Virology 176, 648-651 (1990)
90266484
Draft entry and printed sequence for [1] kindly submitted by
W.Rhode, 15-MAR-1989, for release after publication.
Location/Qualifiers
1. .1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40. .70
/note="stem-loop structure"
103. .975
/note="Orf 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSIVRWCFITLVETEEAAVNVRIESINLVYATVGDEVAPS
TCQHPACGTHKTCPEPLNCIKTVICGNRIHLEPTRGSDENRDYCSKERVLLHGVCP

```

|   |          |  |          |  |       |       |
|---|----------|--|----------|--|-------|-------|
| TRPVKPRRLAQRFAEEDLELGGTRRCVWHGASVETWRWAAENPFPPFYHNMQ<br>LEVISALGEAPADDRITLIWICCRGGGKGKSVFAKYLGLKPDWFTYTCGGTRKDVLYQYIE<br>DCKRNLILDPKRCNLEYLNTALLECVKNRAFSDDKYLEPLSYLGFDDHVLVFPANVLDP<br>VLKTSRDRIKLWNI" | 314..175 | /note="ORF 2"<br>/codon_start=1<br>/protein_id="AAA42895.1"<br>/db_xref="GI:323308"<br>/translation="MTGFTWSRPVPTNRIETTVRRNGCSFSTESRLVLESKGHWDPN<br>DILRLNMSAKNTQADTEDALYTELRWNGDGPKLRSHFHITIGSLACCLURSESQR<br>TAQSSGYADEFTGSGPCLPNLSDSPTGSTHVVPEPTYCTSTRTONEI"<br>complement(422..568)<br>/note="ORF 6"<br>/codon_start=1<br>/protein_id="AAA42896.1"<br>/db_xref="GI:323309"<br>/translation="MEMGTDFQRPILSIPPKLRVQRIFGIRLGLPGGVHQPQQIVGP<br>IVAF" | 639..797 | /note="ORF 3"<br>/codon_start=1<br>/protein_id="AAA42897.1"<br>/db_xref="GI:323310"<br>/translation="MRTERRRREVVRVQISRTQARLVLMHWNNQKGRIVPVHRGPKTK<br>FNPRCTQV"<br>complement(823..987)<br>/note="ORF 5"<br>/codon_start=1<br>/protein_id="AAA42898.1"<br>/db_xref="GI:323311"<br>/translation="MHTNLINPQFVSVPADFQIIRDQIGKYEYHMHVVEPKITKGFFVF<br>RTECPVLNTF"<br>1098..1286<br>/note="ORF 4"<br>/codon_start=1<br>/protein_id="AAA42899.1"<br>/db_xref="GI:323312"<br>/translation="MNRVWGPTIKDSWIITNLLLCLOCTQPLSTPIQVSSLLEKK<br>AASLYLPSICFCAIGRLS" | 332 c | 300 t |
| BASE COUNT  | 336 a    | 323 c  | 332 g    | 300 t  |       |       |
| ORIGIN  |          |  |          |  |       |       |

|                       |                 |                     |                     |                                    |     |
|-----------------------|-----------------|---------------------|---------------------|------------------------------------|-----|
| Query Match           | 100.0%;         | Score 583;          | DB 58;              | Length 1291;                       |     |
| Best Local Similarity | 100.0%;         | Pred. No. 2.3e-161; |                     |                                    |     |
| Matches 583;          | Conservative 0; | Mismatches 0;       | Indels 0;           | Gaps 0;                            |     |
|                       |                 |                     |                     |                                    |     |
| QY                    | 1               | actcgtcctgagtc      | aaaaagccacgattg     | ggcccaacgatttcgtcgtgaggaacctgatgaa | 60  |
|                       |                 |                     |                     |                                    |     |
|                       |                 |                     |                     |                                    |     |
| Db                    | 409             | ACTCGTCCTGGAGTCA    | AAAGGCCACGATTG      | CCCAACGATTTCGTGAGGAACCTGATGAA      | 468 |
|                       |                 |                     |                     |                                    |     |
| QY                    | 61              | ctccgccttgaagacc    | caggcggtacaccgaagat | cgcttgtacacggagcttcggtggaa         | 120 |
|                       |                 |                     |                     |                                    |     |
|                       |                 |                     |                     |                                    |     |
| Db                    | 469             | CTCCGCCTGGAGACCC    | AGCGGATACCGAAGATG   | CGTGTACACGGAGCTTCGGTGGAA           | 528 |
|                       |                 |                     |                     |                                    |     |
| QY                    | 121             | tggacaagatggcgct    | ggaataaccgttccatttc | catatcacaaattggcgagcttggaa         | 180 |
|                       |                 |                     |                     |                                    |     |
|                       |                 |                     |                     |                                    |     |
| Db                    | 529             | TGSACAAGATGGCCGC    | TGAAATCCGTTCCATT    | TCCATATCACAAATTGGCAGCTTGAA         | 588 |
|                       |                 |                     |                     |                                    |     |
| QY                    | 181             | gtgcgtctcgcatcg     | gagaccgcggacgcatgc  | acaaatcctctggatatgcggacga          | 240 |
|                       |                 |                     |                     |                                    |     |
|                       |                 |                     |                     |                                    |     |
| Db                    | 589             | GTGCTGTCTGCCGAT     | CGGAGCCAGCGGACGAT   | CGCAATCTCTTGGATATCGGACGA           | 648 |
|                       |                 |                     |                     |                                    |     |
| QY                    | 241             | gacggaggagacgg      | gaagtccgctgtttgcca  | aatatctcggactcaagccgactggttc       | 300 |
|                       |                 |                     |                     |                                    |     |
|                       |                 |                     |                     |                                    |     |
| Db                    | 649             | GACCGAGGACGCG       | GAAGTCCGTTTGCCAAAT  | ATCTCGACTCAAGCCGACTGGTTTC          | 708 |
|                       |                 |                     |                     |                                    |     |
| QY                    | 301             | tacacatgtgttgg      | aaccagaaaggaogtatt  | gtaccagttacatcagagaccacaaacga      | 360 |
|                       |                 |                     |                     |                                    |     |
|                       |                 |                     |                     |                                    |     |
| Db                    | 709             | TACACATGTGTGG       | AACAGAAAGACGTATT    | GTACCAGTACATCGAGGACCCAAACGA        | 768 |
|                       |                 |                     |                     |                                    |     |
| QY                    | 361             | aatttaatcctcgat     | gtaccgaagtgtaatt    | tagagtatttaaattatgcctgttgaag       | 420 |

```

Db 769 AATTTAAATCCTCGATGTAACCCAGGTGTAATTTACAGTATTTAAATATGACCCCTGTAGAA 828
QY 421 tctgttaagaacaggaattcagtcctggaacaaatcacgaaccccttagttatcttggttc 480
Db 829 TGTGTTAAGACAGGGCAATTCAGTTCGCGACAAATACGAACCCCTTAGTTATCTTGGTTC 888
QY 481 gacctgtgactgtactgtatttgcgaatgctcgtcgtgatttgaataatcacgag 540
Db 889 GACCATGTGCATGTACTCGTATTTGCCAATGCTCTGCTGCTGATTTATTGAAATACAGCAG 948
QY 541 gacagaataaaactgtgaatatttaagaatgtgtcatctaa 583
Db 949 GACAGAATAAAACTGTGAATATTTAAAGTAGTGTGTCATCTAA 991

RESULT 2
LOCUS NV5966 1004 bp DNA circular VRL 13-JAN-1999
DEFINITION faba bean necrotic yellows virus C9 gene.
ACCESSION AJ005966
VERSION AJ005966.1 GI:3550532
KEYWORDS C9 gene; component 9; putative; rep protein; replication associated
protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus
VIRUSES: ssDNA viruses; Nanovirus.
KATUL, L.
1 (bases 1 to 1004)
Direct Submission
Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und
Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und
Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
2 (bases 1 to 1004)
Katul, L., Timchenko, T., Gronenborn, B. and Vetter, H.J.
Ten distinct circular ssDNA components, four of which encode
putative replication-associated proteins, are associated with the
faba bean necrotic yellows virus genome
J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
99094637
FEATURES
source
1..1004
/organism="faba bean necrotic yellows virus"
/virus="Nanovirus"
/isolate="SV292-88"
/db_xref="taxon:59817"
73..918
/gene="C9"
73..918
/gene="C9"
/function="putative replication associated (rep) protein"
/codon_start=1
/product="component 9"
/protein_id="CAA06789.1"
/db_xref="GI:3550533"
/db_xref="SPTREMBL:O91252"
/translation="MSAVNNVFTLNFAGEVPLVLSFDERVOYAVQWHERVNHDIQCVI
QLKKAMNIVKNIIGGNPHLEKMGSIIEASAYAQKEESRVAGWSYGLLKGSH
RKIMELLDPENLEEPQYRRAMASMDSEKRLAEDEGFPYMFYSWQETVLGLIE
EPNDRTIIWYGPNGEKGSGQFLGLKKDLYLPQGTQDMTYMLKPNKANYMD
IPRCSNYLNYQFMELIKNTIYSYKPEVPGCLINKIHVILANVLPDYEKISDRI
KIIYC"
BASE COUNT 352 a 159 c 226 g 267 t
ORIGIN
Query Match 22.6%; Score 132; DB 59; Length 1004;
Best Local Similarity 54.08; Pred. No. 4.5e-28;
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
QY 68 tgaagaccacagcgagataccagagtcggttacacgagcttcggtggaatggacaa 127
Db 419 TGGGAAGACCCAGAAATACAGAGAGCTATGCTTGGTCCGCCATGACGAACTCTCGA 478

```

```

QY 128 gatggcgctgaaatccgttcccatcacaatcgaagcttgcagctgagctgctgt 187
Db 479 AGCTTGGCGAAGAGAGAGGCTTTCCCTATATGTTTACAGTGGCAAGAAACAGTGTGG 538
QY 188 ctgcgattgagagccagcgagcagatcgacacatccctctggtatcgcgagcagcgag 247
Db 539 GCCTATTAGAAAGAGAGCCCAATGACCGTACTATTATTTCGGTCTACGGCCCAATGGTA 598
QY 248 gagcggaagtcgctgttgcgaatattcgaactcaagcccgactggttctacacat 307
Db 599 ATGAAGGAAATACAGTTCGTTGTAATTCCTGGCATTTAAAAAAGATTACCTTTATTAC 658
QY 308 gtgtggaaccagagagcgattgtaccagtcacatcgagcccaacaaatttaa 367
Db 659 CTGGAGGTAAACCCAGATATGACATATATGATTAATGAAAAATCCAAAGGCAATGTTG 718
QY 368 tctcgtgtaccaggtgtaattagatttatttaattatccctgttagaattgttta 427
Db 719 TGATGGATATTCCTCGTGTGAATTCGAATTTAAATTTACCAATTTATGGAATTAATA 778
QY 428 agaacagggcattcagttcggaacaaatcgaaccccttagttatcttgggttcgaccatg 487
Db 779 AATAGAACCATATATAGTTATAAATATGAACCGTTCGATATTAATAATAAAAA 838
QY 488 tgcattctctgtatttgcgaatgctcgtcgtgatttattgaaatcagcagggacagaa 547
Db 839 TACATGTAATTTGATTAGCTAATGATTTCCTCGATTATTCCTGATTAAGAAAAATTAGTCAGGATAGAA 898
QY 548 taaactgtggaattttaa 567
Db 899 TTAATAATAATTTATTGTTAA 918

RESULT 3
LOCUS FBEL32187 1007 bp DNA circular VRL 16-SEP-2000
DEFINITION faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian
EVI-93.
ACCESSION AJ132187
VERSION AJ132187.1 GI:4995171
KEYWORDS C9-Eg gene; rep protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus
VIRUSES: ssDNA viruses; Nanovirus.
1 (bases 1 to 1007)
KATUL, L., TIMCHENKO, T., GRONENBORN, B. and VETTER, H.J.
Ten distinct circular ssDNA components, four of which encode
putative replication-associated proteins, are associated with the
faba bean necrotic yellows virus genome
J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
99094637
JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 1007)
AUTHORS Timchenko, T., de Kouchkovsky, F., Katul, L., David, C., Vetter, H.J.
and Gronenborn, B.
TITLE A single rep protein initiates replication of multiple genome
components of faba bean necrotic yellows virus, a single-stranded
DNA virus of plants
J. Virol. 73 (12), 10173-10182 (1999)
20027244
REFERENCE 3 (bases 1 to 1007)
AUTHORS Katul, L.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie,
Mikrobiol., u. biol. Sicherheit, Biologische Bundesanstalt f. Land-
u. Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
FEATURES
source
1..1007
/organism="faba bean necrotic yellows virus"
/virus="Nanovirus"
/isolate="Egyptian EVI-93"
/db_xref="taxon:59817"
/lab_host="Vicia faba"

```

Thu Oct 25 13:08:07 2001

us-09-462-955-1\_copy\_409\_991.rge

```

stem_loop      1. .39
gene           73. .1003
              /gene="C9-Eg"
CDS            73. .918
              /gene="C9-Eg"
              /function="replication-associated protein"
              /note="32.53 kDa"
              /codon_start=1
              /evidence=experimental
              /product="Rep protein"
              /protein_id="CAB4027.1"
              /db_xref="GI:4995172"
              /db_xref="SP:SPRMBL:Q9WIK2"
              /translation="MSAVNNVFTLNFAGVPLSDERVOYAVWQHERVNHDDHGGVI
              QLKMKMVKNIIGGNPHLEKMGSIIEASAYAKESRVAGPWSGELLKGGSHK
              KRLMKLPDENLEPQKYRRAMASAMDESRLKLAEGGFPYTYLSWQETVLGLLEE
              EPNDRIITWYGPNGNEGSKQFGKFLGKKDYLYLPGGKTQDMTLMKPNKANYMD
              IPRCSEYLNQFMELIKNRTIFSYKYPEVPGCIINNKIHIIVLANVLPDYEKISQDRI
              KIIYC"
TATA_signal    939. .944
              /gene="C9-Eg"
polya_signal   998. .1003
              /gene="C9-Eg"
BASE COUNT    353 a 161 c 227 g 266 t
ORIGIN
1
22.6%; Score 132; DB 58; Length 1007;
Best Local Similarity 54.0%; Pred. No. 4.5e-28;
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
QY 68 tgaagagacccagcgatccaggaagatcggtgtgtacagcgagcttcgtaagtgacaa 127
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 TGAAGAAGACCCAGAAATACAGAAAGAGCGATGGCTTGCTCGCCATGACGAATCTCGGA 478
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 gatggccctgaaatccgttccattccatatacaaatggcagcttgaaagtgcgt 187
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 AGCTTCTGAAGAGGAGCGCTTCCCTATACGCTTTACAGCTGGCAGAAACAGTGTGG 538
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 188 ctgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 247
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 GCCTATTAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 gaacggagagtcggttgcgaatctcgcgactcaagccgagctgttctacacat 307
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 ACCAAGGAAATACAGTTTGGTAATTCCTGGGATTAAGAAAGATTACCTTTATTAC 658
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 gtggtgaacaggaagcgatgtattgtaccaglacatcagagagagagagagagagag 367
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 659 CTGGAGTAAACCCAGATATGACATATATGTAATGAAATATCAAGGCAATGTTG 718
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 tctcgtgtaccaggtgttaattagaattattaaattatgcctgttagaattgttta 427
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 719 TGATGGATATCTCGTGTGAATTCGAATATTTAAATATCAATTAAGCAATTAATA 778
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 agaacaggcattcagttcggacaaatacagaccccttagttatcttgggttcgaccatg 487
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 779 AAAATAGAACCATATTTAGTTTAAATATGACACAGTTCGATGATTAATAATAATAAAA 838
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 tgcgtactcgtatttggcaatgtcctgcctgattatttgaataatcagcagggacagaa 547
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 839 TACATGTAATTGTATTAGCTTAATGTTATGCTGATTGAAATTAATGTCAGGACAGAA 898
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 taaactgtggaattattaa 567
    ||||| || |||||
Db 899 TAAATAAATTTATTGTTAA 918
    ||||| || |||||
RESULT 4
AR063452 AR063452 1096 bp DNA PAT 29-SEP-1999
LOCUS Sequence 4 from patent US 5846705.
DEFINITION
ACCESSION AR063452

```

```

VERSION AR063452.1 GI:5992760
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1096)
AUTHORS Wu,R., You,L. and Soong,T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana
       bunchy top virus and method for detection of banana bunchy top
       virus
JOURNAL Patent: US 5846705-A 4 08-DEC-1998;
FEATURES Location/Qualifiers
           1. .1096
           /organism="unknown"
BASE COUNT 347 a 231 c 244 g 274 t
ORIGIN
1
20.0%; Score 116.4; DB 9; Length 1096;
Best Local Similarity 52.8%; Pred. No. 1.8e-23;
Matches 305; Conservative 0; Mismatches 261; Indels 12; Gaps 2;
QY 6 tctcgtgagtcacaaaggccagcagattggcccaacgatttgcaggagaccctgatgaactcgc 65
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 373 TCCTGGTTTCGAAGAAGCGCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 432
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 66 cctggaagacccagcgagataccgaagatcggttgtacacggagcttcggtggaatggac 125
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 GATGGACATCCATCCAAAGTATCGCAGATGCTTGGCAGTGGAAATCAATTAAGATGCCAG 492
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 126 aagatggccgctgaaaaatccgttccattccatatacaaatggcagcttgaagtgc 185
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 AATTAATTCGGAATGGGTTCACGAACATA-----AAAGAATGGCAAAATAAATTAAT 543
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 gtctcgtcggagagccagcgagcgtatcgacacatcctcttgatgcggagcagacgg 245
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 TCAACACATCGAAGTGTCTCTGATGATCGAAGTATCATCTGGGTATACGCTGCCAAGCG 603
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 246 aggaagacggagtcggtgttggcaaatatctcggactcaagcccgactgggttctacac 305
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 604 AGCGAAGGAAAGTCAACCTTCGCAAGATATCTATCATTTAAACCCGGATGGGATATAT 663
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 306 atggtggaacagaaagagcgtattgtaccagtcagagcagaccccaaacgaaattt 365
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 664 CAACGCTGGAAGACAGCTCGGATATGATGCATCATACAGATGGATCCTGATATCATTTG 723
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 366 atctcctcgtatcccgagtgtaatttagagttatttaattatgcctgttagaattgt 425
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 724 GATTATTGATATCCCGAGAGTCAATTCAGATTATCTGAATTTATGCGTTATAGAACAAAT 783
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 426 taagaacagggcattcagttcggacaaatacgaaccc--ccttagttatcttgggttcga 482
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 784 TAAGAATAGAGTTTAAATAAATACAAATATACGAACCATGTGTGATTAGAAAGATGGACA 843
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 483 ccatgctcgtactcgtatttggcaatgtcctgcctgattatttgaataatcagcagga 542
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 844 AATATGCTCAATGTAATTTATGTCGAATGTTGTCCTGATTATTTGTAATAATTCAGAAGA 903
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 543 cagaataaaactgtggaatatttaaaagtatgtgtcatc 580
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 904 TAGAATAAATAAATAATTAATTTGTTGAGAAAGGAAACTTC 941
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 5
BYTV2 1096 bp DNA VRL 30-OCT-1994
LOCUS Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes,
DEFINITION complete cds.
ACCESSION I32167
VERSION I32167.1 GI:520791
KEYWORDS stem loop.
SOURCE Banana bunchy top virus DNA.
ORGANISM Banana bunchy top virus

```



Viruses; ssDNA viruses; Circoviridae.  
 1 (bases 1 to 1096)  
 Wu, R.-Y., You, L.-R. and Soong, T.-S.  
 Nucleotide sequences of two circular single-stranded DNAs  
 associated with banana bunchy top virus  
 Unpublished (1994)  
 JOURNAL  
 FEATURES  
 Location/Qualifiers  
 1..1096  
 /organism="Banana bunchy top virus"  
 /db\_xref="taxon:12585"  
 1..7  
 TATA\_signal  
 stem\_loop  
 71..928  
 /gene="V2"  
 71..928  
 /gene="V2"  
 /codon\_start=1  
 /protein\_id="AA51426.1"  
 /db\_xref="GI:520792"  
 /translation="MSSFKWCFTLNYSSAAREDFLALLKEEDVHYVVGDEVAPATG  
 QKHLQYLRLKKSIRLGLKKYGRSHWEIAKGSDEONRRYCSKEILVLELGPVVP  
 GSKRRLDLRRFSPLEELKMDPSKYRCLAVESIKDARKINSEWHLEKEMONKLIQH  
 IEGVDPDRSLINVGPNSEGEKSTFARLSLKGWGIYNGKTSIDMHIIITPMDPNHW  
 IIDIPRSHDYLNVGVIDQIKNRVLINTKYEPVIRKDGQNVHVIVMANVLDPYCKLS  
 ERIKILNC  
 85..309  
 /gene="V1"  
 85..309  
 /gene="V1"  
 /codon\_start=1  
 /protein\_id="AA51427.1"  
 /db\_xref="GI:520793"  
 /translation="MVLHSEFLRLSGARRLSRSEGGCSLLCRRRRSRSGHRPEAPP  
 GISPEKINSRRIKEVWLPLSLGDCERK"  
 complement(544..699)  
 /gene="C2"  
 complement(544..699)  
 /gene="C2"  
 /codon\_start=1  
 /protein\_id="AA51428.1"  
 /db\_xref="GI:520794"  
 /translation="MMCLISDVFPPLIYPHPGNDRLAKVDFPSPPLGPTQMILRS  
 SGTFSMC"  
 complement(883..1059)  
 /gene="C1"  
 complement(883..1059)  
 /gene="C1"  
 /codon\_start=1  
 /protein\_id="AA51429.1"  
 /db\_xref="GI:520795"  
 /translation="MGLRPNKRFINYNDFLHLMIACSASRYVSGTCTFLIACGGSFLS  
 QQLTIFILSSEILQ"  
 347 a 231 c 244 g 274 t  
 BASE COUNT  
 ORIGIN

Query Match  
 Best Local Similarity 20.8%; Score 116.4; DB 58; Length 1096;  
 Matches 305; Conservative 0; Mismatches 261; Indels 12; Gaps 2;  
 QY 6 tcttgagtcacaaagccacgattggcccaacgatttgcgtgaggaacctgatgaactccg 65  
 DB 373 TCCTGGTTCGAAGAGCGCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 432  
 QY 66 cctggagaccagcggcgatccacgaatgcgtgtgtacacgagcttcggtggaatggac 125  
 DB 433 GATGACCATCCATCCCAAGTATCCGATGCTTGGCAGTGGGAATCAATTAAGATGCCAG 492  
 QY 126 aagatggcgctgaaatccgttccattccatatacaaatggcagcttgaagtgc 185  
 DB 493 AATTAATTCGAATGGGTTCACGAACATA-----AAAGATGCCAAATAAATAAT 543  
 QY 186 gtctgcgatcgagagccagcgacgatcgacacaaatctctctggtatgcgagcagcg 245

Db 544 TCACACATCGAAGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGGTCCCAACGG 603  
 QY 246 agggacgggaagtcggtgtttgccaataatctcggaactcaagcccgactgggtctacac 305  
 Db 604 AGGCGAAGGAAGTCAACCTTCGCAAGATATCTATCATTAATAAACCCGGATGATAT 663  
 QY 306 atgtggtggaccagaagaagcgtattgtaccagtagcatcgagaccccaacgaattt 365  
 Db 664 CAACGGTGGGAAGACGTCGGATATGATGCACATCATAAACGATGGATCTGATAATCAT 723  
 QY 366 aatcctcgatgacccaggtgtaatttagagatttataaattatgcctgttagaattgt 425  
 Db 724 GATTATTGATATCCCAAGAAGTCAATTCAGATTATCTGAATTATGGCGTTATAGAACA 783  
 QY 426 tagaacagggcattcaggttcggacaaaatacgaacc---ccttagttatcttgggttga 482  
 Db 784 TAGAATAGAGTTTTATAATAATACAAAATACGAACCATGTGTGATTAGAAAAAGATGG 843  
 QY 483 ccattgcatgtactgatttcccaatgctcctgctgattatttgaataatcagcagga 542  
 Db 844 AATGTCCATGTAATTGTTATGCAAAATGTGTGCTGATTATTGTAAATTTTCAGAA 903  
 QY 543 cagaataaaactggaataatttaaagtagtgcac 580  
 Db 904 TAGAATAAAATAAATTAAATTGTTGAGAAAGAACTTC 941  
 RESULT 6  
 AR063453  
 LOCUS AR063453 1091 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 5 from patent US 5846705.  
 ACCESSION AR063453  
 VERSION AR063453.1 GI:5992761  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 1091)  
 Wu, R., You, L. and Soong, T.  
 Nucleotide sequence of two circular ssDNA associated with banana  
 bunchy top virus and method for detection of banana bunchy top  
 virus  
 JOURNAL Patent: US 5846705-A 5 08-DEC-1998;  
 FEATURES Location/Qualifiers  
 1..1091  
 BASE COUNT 343 a 224 c 248 g 276 t  
 ORIGIN

Query Match  
 Best Local Similarity 19.78%; Score 114.8; DB 9; Length 1091;  
 Matches 304; Conservative 0; Mismatches 262; Indels 12; Gaps 2;  
 QY 6 tcttgagtcacaaagccacgattggcccaacgatttgcgtgaggaacctgatgaactccg 65  
 Db 368 TCCTGGTTCGAAGAGCGCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 427  
 QY 66 cctggagaccagcggcgatccacgaatgcgtgtgtacacgagcttcggtggaatggac 125  
 Db 428 GATGACCATCCATCCCAAGTATCGAGATGCTTGGCAGTGGGAATCAATTAAGATGCCAG 487  
 QY 126 aagatggcgctgaaatccgttccattccatatacaaatggcagcttgaagtgc 185  
 Db 488 AATTAATTCGAATGGGTTCACGAACATA-----AAAGATGCCAAATAAATAAT 538  
 QY 186 gtctgcgatcgagagccagcgacgatcgacacaaatctctctggtatgcgagcagcg 245  
 Db 539 TCACACATCCCAAGTGTTCCTGATGATCGAAGTATCATCTGGGTATACGGTCTCGCCGG 598  
 QY 246 agggacgggaagtcggtgtttgccaataatctcggaactcaagcccgactgggtctacac 305

Best Local Similarity 56.9%; Pred. No. 7.2e-23;  
Matches 230; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

599 AGGCGAAGGAAAGTCAACCTTCGCAAGATATCTATCATTAACCTGGATGGGGATATAT 658  
306 atgtggggaacaggaagcattgtaccagtcacatcagagcccaacgaattt 365  
659 CAACGGTGGAAAGACGTCGATATGATGCACATCAACGATGGATCCCTGATAATCAT 718  
366 aatcctcgatgaccaggtgtgaatttagagattttaaattatgacctgttagaatgtgt 425  
719 GATTATTGATATCCCAAGAGTCAATTCAGATATCTGATTAATGATGGCTTATAGAACAA 778  
426 taagaacagggcaltcagttcggaacaatacgaacc---ccttagttatcttgggttga 482  
779 TAAGAATAAGATGTTTAATAATACAAATACGAACCATGTCTGATTAGAAAAGATGGACA 838  
483 ccattgcatgactcgtattgccaattgctcgtcctgctgattattgaaatacagcgga 542  
839 AATGTCATGTAATGTTATGCAAAATGCTTGCCTGATTATTGTAATAATTTTCAAGA 898  
543 cagaataaacgtggaattattaaagtattgtgtc 580  
899 TAGAATAAAATAATTAATTTGTTGAGAAAGGAAACTTC 936

RESULT 7  
AF216222 AF216222 1095 bp DNA VRL 08-MAR-2001  
LOCUS Banana buncy top virus satellite S2 replication initiation protein  
DEFINITION (ORF V1) gene, complete cds.  
ACCESSION AF216222  
VERSION AF216222.1 GI:12004327  
KEYWORDS  
SOURCE Banana buncy top virus.  
ORGANISM Banana buncy top virus  
REFERENCES 1 (bases 1 to 1095)  
AUTHORS Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.  
TITLE Additional rep-encoding HDAs associated with banana buncy top virus  
JOURNAL Arch. Virol. 146 (1), 71-86 (2001)  
MEDLINE 21163975  
REFERENCE 2 (bases 1 to 1095)  
AUTHORS Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.  
TITLE Direct Submission  
JOURNAL Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia

FEATURES  
source  
1. .1095  
/organism="Banana buncy top virus"  
/isolate="Taiwan"  
/db\_xref="taxon:12585"  
/note="satellite S2"  
64..921  
/gene="ORF V1"  
64..921  
/gene="ORF V1"  
64..921  
/gene="ORF V1"  
/codon\_start=1  
/product="replication initiation protein"  
/protein\_id="AAG44004.1"  
/db\_xref="GI:12004328"  
/translation="MSSFKWCTFLNYSSAAREDFLALLKEELNYAVGVGDEVPSTG  
RKLOGLYSLKKSILKGLKRYSSKAHWAARGTDEQNRGYSKETLVLELGPVVP  
GSKRLLEFRSPKLEPMEDPSKYRCLAEVSNARKNSWHELREWNKLIQH  
IEGVPPDRSTLWVYGPNGEGKSTFARYLSLKPWGTYNGGKTSMDHIIITMDPDNH  
IIDIPIRSHDLYNGVIEQIKRVLINTKYECVPIKDGQNVHIVMANVLPDYCKIS  
EDRIKIINC"  
BASE COUNT 347 a 219 c 252 g 277 t  
ORIGIN

Query Match 19.6%; Score 114.4; DB 58; Length 1095;

167 attggcagcttgaaagtgcgtctgcgacgcgagagccagcgacacgcacaatctctt 226  
518 ANTGGCAAAATAAATTAATTCACACATCGAAGGTCTCTGATGATCGAAGTATCATCT 577  
227 ggatcagcgacgagcagcgagagcggaagtccgtgtttgccaaatatctcgactca 286  
578 GGGTATACGGTCCCAACGGAGGCGAAGAAAGTCAACCTTCGCAAGATATCTATCA 637  
287 agcccgactgggttctacacatgtgtgggaaccagaaagagcgtattgtaccagtcac 346  
638 AACCTGGATGGGATATATCAACGGTGGAAAGACGTCGGATATGATGCACATCATCA 597  
347 aggacccaaacgaaatttaattcctcgatgaccaggtgttaatttagagattttaa 406  
698 TGGATCCTTGATTAATCATGTTGGATTATGATATCCCAAGAGTCAATTCAGATT 757  
407 atgcccgttagaattgtttaagaacagggcattcaggttcggacaatacgaacc---cc 463  
758 ATGGCTTATAGAACAAATTAAGAATAGAGTTTTTAATAACAAATACGAACCATGTG 817  
464 ttatttctctgggttcgaccatgctgactcgtattgccaattgctcgtcgtgatt 523  
818 TCATTAGAAAAGATGGACAAATGTCCTGATTAATGTTATGCGCAATGCTTGCCTG 877  
524 atttgaatacagcagggagcagaaataaaactgtggaattatttaa 567  
878 ATTGTAAATTTTCAGAGATAGATAAAATAAATAATTAATTTGTTGA 921

RESULT 8  
BBU12586 BBU12586 1111 bp DNA circular VRL 01-FEB-1995  
LOCUS Banana buncy top virus DNA III ORF V1 and ORF C1 genes, complete  
DEFINITION cds.  
ACCESSION U12586  
VERSION U12586.1 GI:642390  
KEYWORDS  
SOURCE Banana buncy top virus.  
ORGANISM Banana buncy top virus  
REFERENCES 1 (bases 1 to 1111)  
AUTHORS Wu,R.-I. and You,L.-R.  
TITLE Nucleotide sequences of DNA III and DNA IV associated with banana buncy top virus and their relation to other closely related virus DNAs

FEATURES  
source  
1. .1111  
/organism="Banana buncy top virus"  
/db\_xref="taxon:12585"  
/clone="H-4"  
1..8  
TATA\_signal  
stem\_loop  
15..49  
CDS  
79..933  
/note="ORF V1"  
/codon\_start=1  
/protein\_id="AAA61875.1"  
/db\_xref="GI:642391"  
/translation="MSSFKWCTFLNYSSAAREDFLALLKEELNYAVGVGDEVPSTG  
OKHLOGLYSLKKSILKGLKRYSSKAHWAARGTDEQNRGYSKETLVLELGPASQ  
GNNRRLSEMVSPERMRIEQEYIHYRTSVKLLKRFKEFHPCLDRPQIOLTER  
IDEPDRTIFWYGPNGEGKSTYAKLMKKDWFTYTGKKNELLSYVDEGSEKHI  
VDFIPRCNODLYNDVIEALKDRVIESTKYPIKLVELNNHIVMANFMPECKISE  
DRKIIVC"



[illegible]



```

FEATURES             Location/Qualifiers
     source            1..1106
                        /organism="unknown"
BASE COUNT           335 a   225 c   257 g   289 t
ORIGIN
169 tggcagctgaagtcgtctgcgacgcgagagcagcgagcgcgcacacacccctctgg 228
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
524 TGGCAGATTCAATACCGAGGCAATACCGAGGCAATACCGAGGCAATACCGAGGCAAT 583
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
529 atatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
584 GTCATATGTCCTTATGTTAATGAGGTTAATCAACATATGCGAAGTCACATCAAGAAG 643
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
589 ccgactggttccacacatgtgtggaacccgagagcagcagcagcagcagcagcagcag 348
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
644 GATTGTTCTACACCGAGGCAATACCGAGGCAATACCGAGGCAATACCGAGGCAAT 703
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 gaccccaaaacgaatttaacctcgcgacgcgagcagcagcagcagcagcagcagcagcag 408
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
704 GGATCTGACAGCATATAGTATTTGATATTCCTCGTTGATATCAGGATTTTAAATAT 763
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 gccctgttagaagtgtttaagaacccgagcagcagcagcagcagcagcagcagcagcag 468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 GATGTAATAGAGGCAATTAAGGATAGGTTATAGAGAGTACTAAATACAAACCCATAAG 823
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 tatcttggttgcacacatgtgtggaacccgagagcagcagcagcagcagcagcagcag 528
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
824 ATAGTTGAATAGGTTAATGAGGTTAATCAACATATGCGAAGTCACATCAAGAAG 926
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Query Match      17.5%; Score 102.2; DB 9; Length 1106;
Best Local Similarity 53.3%; Pred. No. 2.9e-19;
Matches 215; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 169 tggcagctgaagtcgtctgcgacgcgagagcagcgagcgcgcacacacccctctgg 228
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 TGGCAGATTCAATACCGAGGCAATACCGAGGCAATACCGAGGCAATACCGAGGCAAT 583
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 229 atatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 GTCATATGTCCTTATGTTAATGAGGTTAATCAACATATGCGAAGTCACATCAAGAAG 643
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 289 ccgactggttccacacatgtgtggaacccgagagcagcagcagcagcagcagcagcag 348
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 644 GATTGTTCTACACCGAGGCAATACCGAGGCAATACCGAGGCAATACCGAGGCAAT 703
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 349 gaccccaaaacgaatttaacctcgcgacgcgagcagcagcagcagcagcagcagcagcag 408
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 704 GGATCTGACAGCATATAGTATTTGATATTCCTCGTTGATATCAGGATTTTAAATAT 763
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 409 gccctgttagaagtgtttaagaacccgagcagcagcagcagcagcagcagcagcagcag 468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 764 GATGTAATAGAGGCAATTAAGGATAGGTTATAGAGAGTACTAAATACAAACCCATAAG 823
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 469 tatcttggttgcacacatgtgtggaacccgagagcagcagcagcagcagcagcagcag 528
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 824 ATAGTTGAATAGGTTAATGAGGTTAATCAACATATGCGAAGTCACATCAAGAAG 926
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 529 aaaaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 571
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 AAAATCTCGAAGATCGAATAAATATCAATTTATTTGCTGAAGAA 926
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
NV5964
LOCUS             NV5964             1014 bp      DNA      circular      VRL
DEFINITION       faba bean necrotic yellows virus C7 gene.
ACCESSION       AJ005964
VERSION         AJ005964.1   GI:3550512
KEYWORDS        C7 gene; component 7; putative; rep protein; replication associated

```



us-09-462-955-1\_copy\_409\_991.rge

Thu Oct 25 13:08:07 2001

-----



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:35 ; Search time 5701.85 Seconds  
(without alignments)  
2259.729 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_383\_1215

Perfect score: 833

Sequence: 1 ggggtcttcagcacgga.....accatccagtgagtagct 833

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_om.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_pl1.\*
- 13: gb\_pl2.\*
- 14: gb\_pl3.\*
- 15: gb\_pl4.\*
- 16: gb\_bal.\*
- 17: gb\_ba2.\*
- 18: gb\_fun.\*
- 19: em\_htgo\_hum.\*
- 20: em\_htgo\_inv.\*
- 21: em\_htgo\_rod.\*
- 22: em\_htg\_hum1.\*
- 23: em\_htg\_hum2.\*
- 24: em\_htg\_hum3.\*
- 25: em\_htg\_hum4.\*
- 26: em\_htg\_hum5.\*
- 27: em\_htg\_hum6.\*
- 28: em\_htg\_hum7.\*
- 29: em\_htg\_hum8.\*
- 30: em\_htg\_inv1.\*
- 31: em\_htg\_inv2.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_rod.\*
- 34: em\_hum1.\*
- 35: em\_hum2.\*
- 36: em\_hum3.\*
- 37: em\_hum4.\*
- 38: em\_hum5.\*
- 39: em\_hum6.\*
- 40: em\_hum7.\*
- 41: em\_in.\*
- 42: em\_om.\*
- 43: em\_or.\*

44: em\_ov.\*

45: em\_pat.\*

46: em\_ph.\*

47: em\_pl.\*

48: em\_ro.\*

49: em\_sts.\*

50: em\_sy.\*

51: em\_un.\*

52: em\_vi.\*

53: gb\_sts1.\*

54: gb\_sts2.\*

55: gb\_sts3.\*

56: gb\_sy.\*

57: gb\_un.\*

58: gb\_vil.\*

59: gb\_vil2.\*

60: gb\_vil3.\*

61: gb\_vil4.\*

62: gb\_vil5.\*

63: gb\_vil6.\*

64: gb\_vil7.\*

65: gb\_vil8.\*

66: gb\_vil9.\*

67: gb\_vil10.\*

68: gb\_vil11.\*

69: gb\_vil12.\*

70: gb\_vil13.\*

71: gb\_vil14.\*

72: gb\_vil15.\*

73: gb\_vil16.\*

74: gb\_vil17.\*

75: gb\_vil18.\*

76: gb\_vil19.\*

77: gb\_vil20.\*

78: gb\_vil21.\*

79: gb\_vil22.\*

80: gb\_vil23.\*

81: gb\_vil24.\*

82: gb\_vil25.\*

83: gb\_vil26.\*

84: gb\_vil27.\*

85: gb\_vil28.\*

86: gb\_vil29.\*

87: gb\_vil30.\*

88: gb\_vil31.\*

89: gb\_vil32.\*

90: gb\_vil33.\*

91: gb\_vil34.\*

92: gb\_vil35.\*

93: gb\_vil36.\*

94: gb\_vil37.\*

95: gb\_vil38.\*

96: gb\_vil39.\*

97: gb\_vil40.\*

98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|-------------|
| 1          | 833   | 100.0         | 1291   | 58    | CFDCG       |
| 2          | 132   | 15.8          | 1004   | 59    | NYV5966     |
| 3          | 132   | 15.8          | 1007   | 58    | FBE132187   |
| 4          | 120.6 | 14.5          | 1096   | 9     | AR063452    |
| 5          | 120.6 | 14.5          | 1096   | 58    | BYTV2       |
| 6          | 117   | 14.0          | 1091   | 9     | AR063453    |
| 7          | 115.6 | 13.9          | 1095   | 58    | AF216222    |
| 8          | 108.6 | 13.0          | 1111   | 58    | BBU12586    |

M29963 Coconut fol  
AJ005966 faba bean  
AJ132187 faba bean  
AR063452 Sequence  
L32167 Banana bunc  
AR063453 Sequence  
AF216222 Banana bu  
U12586 Banana bunc

## ALIGNMENTS

| RESULT | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 |  |
|--------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|--------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|



DNA virus of plants  
J. Virol. 73 (12), 10173-10182 (1999)  
20027244

JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL

Direct Submission  
Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Land-Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. Land-u. Forstwirtschaft, Messeweg 11 -12, 38104 Braunschweig, GERMANY

Location/Qualifiers

1..1007 /organism="faba bean necrotic yellows virus"  
/virion  
/isolate="Egyptian EVI-93"  
/db\_xref="taxon:59817"  
/lab\_host="Vicia faba"

stem\_loop gene 1..39  
73..1003 /gene="C9-Eg"

CDS 73..918  
/gene="C9-Eg"  
/function="replication-associated protein"  
/note="32.53 kDa"  
/codon\_start=1  
/evidence=experimental  
/product="Rep protein"  
/protein\_id="CAB44027.1"  
/db\_xref="GI:4995172"  
/db\_xref="SPTREMBL:Q9WIK2"  
/translation="MSAVNVFTLNFAGEVPLSFDERVOYAVMHERVNHDIQVI OLKKAKMTVNKIIGGNPHLEKMKSLSEASAYAQKEESRVAGPWSYGELKKGSHK RKTMLIKDPENLEEPPOKRYRAMAWSAMDESRLKABEGFPPTYLYSWQETVLGLLEE EPNDRILINVGPNNEGKSQRGFGLKDKDYLPGKTDMDTYLMKNPKNAVYMD IPRCNSEYLYNPFMELINKRITFSKYEPVCIIINKHIVILVANLPDYEKISQDRI KIIFYC"

TATA\_signal 939..944 /gene="C9-Eg"  
polyA\_signal 998..1003 /gene="C9-Eg"

BASE COUNT 353 a 161 c 227 g 266 t

ORIGIN

Query Match 15.8%; Score 132; DB 58; Length 1007;  
Best Local Similarity 54.0%; Pred. No. 7.9e-27;  
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 94 tggaagaccggcggtacccaagaatgcgttgtcacaggagcttcggtggaatggacaa 153  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 419 TGGGAAGACCCCGAGAAATACAGACAGCGGTGCCTGGTCCGCCATGACGAATTCGGA 478  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 154 gatggcgcgtgaaaatccgttccccattccatatcaccaattggcagcttgagtgcgt 213  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 479 AGCTTGCTGAAGAAGAGGCTTTCCCTATACGCTTTACAGCTGGCAAGAACAGTGTGG 538  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 214 ctgcgatcgagagccagcgagcatcgacaatccctctgatatcgacgagcggag 273  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 539 GCCTATTAGAAGACAGCCCCCATACCCTATTATTATTGGGTCTACGCCCCCAATGGTA 598  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 274 gagacgggaagtcggtttgttgcacaaatactcgcactcaagcccgcactgggttccacat 333  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 599 ACGAAGAAAATCACAGTTTGGTAAATTCCTGGGATTTAAAAAAGATTACCTTTATTATAC 658  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 334 gtggtggaaccgaaagacgattgttaccagtagtatcgaggagcccaaacgaaatttaa 393  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 659 CTGGAGGTAAAAACCAAGATATGACATATATGTTAAAGAAATCCAAAGCAATGTTG 718  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 394 tcctcgatgtaccaggtgaatttagagtagtatttaaatatgacctgttagaattgttta 453  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 719 TGATGATATTCCTCGTGTANTTCCTGAATTTTAAATTTATCAATTTATGGAATTAATTA 778  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 454 agaacagggcattcagttcggcaaatcagaccccttagttatcttgggttcgaccatg 513  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||



```

Query Match          14.0%; Score 117; DB 9; Length 1091;
Best Local Similarity 52.2%; Pred. No. 1.6e-22;
Matches 314; Conservative 0; Mismatches 275; Indels 12; Gaps 2;

      9   ctcgagcaccgaggtccgcgactgcctcgtagtcaaaaggccacgattggcccacaagtatt 68
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
     345 CTTGAATGGGTACTCGGTGGTTCCTGGTTCGAGAACGCCGAAGCTTCTCGATAGATTCC 404

     69   gctgaggaaacctgatgaactccgcttggaagaccocaggcgatcacgagaatgcgttgtta 128
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    405 AGAGAGAGCCCTGAGGAATTTGAAGATGGGACGATCCATCCAAGTATCGCAGATGCTTGACA 464

    129   cacgagacttcggtggaatggacaagatgggccgctgaaaatccgttccccattcccatat 188
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    465 GTGGAATCAATTAAGAATGCCAAGTAATTCGGAATGGGTTTCACGAACTA----- 515

    189   cacaattggcagcttggaagtgcgtgctgogatcggagagccaggcagatatcgacaatc 248
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    516 AAAGAATGGCAAAAATAAATTAATTCACACATCCGAAGGTGTTCCCTGATGATCGAAGTATC 575

```

|       |            |  |                 |
|-------|------------|--|-----------------|
|       | 249        | ctctgatatcgcagacgacgagcgagcggaagtccgtgtttgccaaatattctcggga           | 308             |
| y     |            |  |                 |
| b     | 576        | ATCTGGGTATTACGGTCTCGGAGGCCGAAGAAGTCACCTTCGCAAGATATCTATCA             | 635             |
|       |            |  |                 |
| y     | 309        | ctcaagcccgcactggttctacacatgtggttggaaccagaagaagacgatgtaccagtag        | 368             |
| b     |            |  |                 |
| y     | 636        | TATAAACCTGGATGGGATATATCAACGGTGGAAGACGTCGGATATGATGCACATCAT            | 695             |
|       |            |  |                 |
| y     | 369        | atcaggagccccaaaacgaataattactcgtatgccaggltgaatttagagtatatta           | 428             |
| b     |            |  |                 |
| y     | 696        | ACGATGGATCTGTGTAATCATTTGGATATTATTGATATCCCACAGAGTCATTCAGATTATCTG      | 755             |
|       |            |  |                 |
| y     | 429        | aattatgccctgttagaatgtttaagaacacagggcattccagttcggacaataatcgaacc-      | 487             |
| b     |            |  |                 |
| y     | 756        | AATTATGGCGTTATAGACAATAAAGAATAAGACTTTTTAAATAAATACAAAATACGAACCA        | 815             |
|       |            |  |                 |
| y     | 488        | --cctagtattcctgggttcgacctatgcatatactgtaatttgcgaatgctcctgcct          | 545             |
| b     |            |  |                 |
| y     | 816        | TGTGTGATTAGAAAAAGATGGCAAAATGTCATGTAATGTTATGGCAAATGTGTGCCT            | 875             |
|       |            |  |                 |
| y     | 546        | gattattgaaatcagcagggacagaataaaactgtggaatatattaaagtatgtgcat           | 605             |
| b     |            |  |                 |
| y     | 876        | GATTATTGTAATAATTTCCAGAGATAGATAAAATAATTAATTGTTTCAGAAAGAAACTT          | 935             |
|       |            |  |                 |
| y     | 606        | c 606  |                 |
| b     | 936        | c 936  |                 |
| <hr/> |            |  |                 |
|       | RESULT     | 7  |                 |
|       | AF216222   | DNA  | VRL 08-MAR-2001 |
|       | LOCUS      | Banana bunchy top virus satellite S2 replication initiation protein  |                 |
|       | DEFINITION | (ORF VL) gene, complete cds.   |                 |
|       | ACCESSION  | AF216222   |                 |
|       | VERSION    | AF216222.1 GI:12004327   |                 |
|       | KEYWORDS   | .  |                 |
|       | SOURCE     | Banana bunchy top virus.   |                 |
|       | ORGANISM   | Banana bunchy top virus.   |                 |
|       | REFERENCE  | Viruses: ssDNA viruses; Circoviridae.                                |                 |
|       | AUTHORS    | 1 (bases 1 to 1095)  |                 |
|       | TITLE      | Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.                    |                 |
|       |            | Additional rep-encoding DNAs associated with banana bunchy top virus |                 |
|       | JOURNAL    | Arch. Virol. 146 (1), 71-86 (2001)                                   |                 |
|       | MEDLINE    | 21163975   |                 |
|       | REFERENCE  | 2 (bases 1 to 1095)  |                 |
|       | AUTHORS    | Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.                    |                 |
|       | TITLE      | Direct Submission  |                 |
|       | JOURNAL    | Submitted (14-DEC-1999) School of Life Sciences, Queensland          |                 |
|       |            | Orf. 4001  |                 |

Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001,

| Query Match           | 13.0%;       | Score 108.6;  | DB 58;          | Length 1111; |
|-----------------------|--------------|---|-----------------|--------------|
| Best Local Similarity | 57.0%;       | Pred. No. 4e-20;  |                 |              |
| Matches 239;          | Conservative | 0;  | Mismatches 174; | Indels 6;    |
| Gaps                  | 2;           |   |                 |              |
| QY                    | 195          | tggcagcttgaagtgctgtctgcgctgagcggagccagcgagcgtatcgcacaaatcctctgg   | 254             |              |
| Db                    | 535          | TGGCAGATTCATTTGACGGAGGCAATGTGACGAGGACCAGATGATCGAACGATCTCTGG       | 594             |              |
| QY                    | 255          | atatcgcgacgagcggagcggagcggggaagtcgcgtgtttgccaaatatctcggactcaag    | 314             |              |
| Db                    | 595          | GTCTATGTGTCGGAATGGTAATGATGGGGAATCAACATATGCCAAGCTCATT---AATGAAG    | 651             |              |
| QY                    | 315          | cccgcactgggtttcacacatgtgttgaaacccaggaaggcgtattgtaccagtcacat---c   | 371             |              |
| Db                    | 652          | AAGCACTGGTCTTACACGAGGTGGGAAGAGGAGAACATACCTGTTCTCTTACGTGGAC        | 711             |              |
| QY                    | 372          | gaggaccaccaaaacgaaatttaatccctcgcagtgcaccacgggtgaatttagagattttaaat | 431             |              |
| Db                    | 712          | GAAGGATCTCAGAAAGCATATTTGTTATTCCTTCGCTGTAATCAGGATTAATTTAAAT        | 771             |              |
| QY                    | 432          | tatgccctgttaagaatgtttaagaaacagggcattcagttcggacaaataacgaacccctt    | 491             |              |

|                       |              |                   |                |             |
|-----------------------|--------------|-------------------|----------------|-------------|
| Query Match           | 13.0%        | Score 108.2       | DB 58          | Length 1127 |
| Best Local Similarity | 56.7%        | Pred. No. 5.2e-20 |                |             |
| Matches 241           | Conservative | 0                 | Mismatches 178 | Indels 6    |
|                       |              |                   |                | Gaps 2      |

|    |     |  |     |
|----|-----|--|-----|
| QY | 195 | tgcgcagtgaagtgctgctgcgatacggagcagcgagcagatcgcaaatcctctgg   | 254 |
|    |     |  |     |
| Db | 535 | TGCGACATTCAATTGACGGAGCCAATTGACGAGGAACCGATGATCGAAGCATCTCTGG | 594 |
|    |     |  |     |
| QY | 255 | atatcggcgcagcagcagcagcagcagtcgctgtttgccaaatatctcgagctctaag | 314 |

GYLOMKKOTLLKMKLELLPAGHLEMARAPKKAIDYCKQKETAIDCPWEYGTWISTGSH  
 KKKLREDEDEEMKLEDPGLYRCLSRVOMTKVREKNSWDYDLRWQDLELLKIEQ  
 EDDRTILWYICPHGEGKSPFAXLTLKEGWNVTAGKATDMLYSYLDPTCHVICID  
 IPKTRKEINYAVIEQIKNRIIINIKTEPCTIRDHGHVHVIYFANYLPDVTRISED  
 RIKIIVC"  
 887..892  
 /gene="Rep"  
 898..903  
 /gene="Rep"  
 BASE COUNT 348 a 179 c 232 g 263 t  
 ORIGIN  
 Query Match 12.7%; Score 105.4; DB 58; Length 1022;  
 Best Local Similarity 51.5%; Pred. No. 3.3e-19;  
 Matches 325; Conservative 0; Mismatches 291; Indels 15; Gaps 3;  
 QY 12 gagcagcaggtcccgactcgtccctggagtcacaaaggccagattgcccacacgatttgc 71  
 Db 330 GAGTATGGGACCTGGATTCTTACAGGTAGTCACAAACGAAAGCTTATGGAAGATTGAT 389  
 QY 72 gaggaacctgataactccgctggaagaccgagcggtataccgaagatcgcttgatcac 131  
 Db 390 GAAGACCCAGAGAGATGACGCTTGAAGATCCAGGATTATACAGCGGATCTTT----- 442  
 QY 132 ggagcttcggtggaatggacaagatgggcccgtgaaatccggtccattcccatatcac 191  
 Db 443 --ATCGGAGTTTCAGATGACTTAAAGTTAGAGAGAGAAATTTCTTGGGATTACGATCTTCGT 500  
 QY 192 aattggcagctgaagtctgtctgcagtcgagcagcgagccagcgagcgcacaaatcctc 251  
 Db 501 CCATGGCAGGATGAATTTGTTGAAGACGATTGAACAGNACCAGATGATCGTACTATTCTG 560  
 QY 252 tgatatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 311  
 Db 561 TGGGTCTATGGCCACATGCGAGCGAGGCGAAAGTGTCTTTGCTAAATACTTAACTTAACTG 620  
 QY 312 aagcccgactggtctacacatggtggaaccagagagagagagagagagagagagagagag 371  
 Db 621 AAGGAAGTTGGTGTATACAGCTGGAGGGAAGAGCAGAGATATGCTTTACAGCTATTCA 680  
 QY 372 gagcaccac 428  
 Db 681 TTGACCCACATGCCACATGCTATGATATACCCAGATGTACTTAAAGAAAGAAATATATT 740  
 QY 429 aattgcccgttgaatggttgaagacagcagcagcagcagcagcagcagcagcagcagc 487  
 Db 741 AATTACGCACTAATTGAGCAAAATTAATAATAGAAATTAATAATTAATAATTAATAATTA 800  
 QY 488 --ccttagttatctgggttcgacacatgctgactgctgactgctgactgctgactgctg 545  
 Db 801 TGCATTATTAGACAGATGCTATATGTCATATGTCATATGTCATATGTCATATGTCATAT 860  
 QY 546 gattatttgaatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 605  
 Db 861 GATGTAACAGGATTAGTGAAGACAGATAAATAAATAAATAAATAAATAAATAAATAAATA 920  
 QY 606 ctaataac 636  
 Db 921 TTAATAACTACGCGAAGCGTATTGGACACGC 951  
 RESULT 11  
 AF216221  
 LOCUS AF216221 1109 bp DNA VRL 08-MAR-2001  
 DEFINITION Banana bunchy top virus satellite S1 replication initiation protein  
 (ORF V1) gene, complete cds.  
 ACCESSION AF216221  
 VERSION AF216221.1 GI:12004325  
 KEYWORDS Banana bunchy top virus.  
 SOURCE Banana bunchy top virus.  
 ORGANISM Viruses; ssDNA viruses; Circoviridae.  
 595 GTCATATGTCGCAATGTAATGAGGGAATCAACATATCGAAGTCAIT---AATGAAG 651  
 QY 315 cccgactggtctacacatggtggaacacagcagcagcagcagcagcagcagcagcagcagc 371  
 Db 652 AAGGACTGTTCTACACAGAGGTGGGAAGAGGAGAACTACTGTTCTCCTACGTGGAC 711  
 QY 372 gagacccac 431  
 Db 712 GAAGGATCTGAGACCATATGTTATTCATATTCCTGCTGTAATCAGGATTTATTTAAAT 771  
 QY 432 tatgcctcttagaattggtttaagacagcagcagcagcagcagcagcagcagcagcagc 491  
 Db 772 TATGATGTTATAGAGCATTAAGAGTAGGTGATAGAGAGTACTATAAATAAATCACTATT 831  
 QY 492 agttatctggttcgacacatgctgctgctgctgctgctgctgctgctgctgctgctgct 551  
 Db 832 AAGTTAGTTGAATGAATAATACATGATTAATGTCATGCTAATTCATGCGCAATTC 891  
 QY 552 ttgaaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 611  
 Db 892 TGTAAATCTCCGAAGAGAGATAAAGATTATTTATTGTTAAATAACACGCTATGACAAT 951  
 QY 612 tacac 616  
 Db 952 CGTAC 956  
 RESULT 10  
 AB009047 1022 bp DNA circular VRL 30-OCT-1998  
 LOCUS Milk vetch dwarf virus genome segment 10 encoding replication  
 DEFINITION Initiation protein, complete sequence.  
 ACCESSION AB009047  
 VERSION AB009047.1 GI:3798737  
 KEYWORDS putative viral replication-initiation protein.  
 SOURCE milk vetch dwarf virus (specific\_host:Astragalus sinicus,  
 lab\_host:Pisum sativum) DNA.  
 ORGANISM  
 Viruses; ssDNA viruses; Nanovirus.  
 REFERENCE 1 (bases 1 to 1022)  
 AUTHORS Sano.Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-NOV-1997) to the DDBJ/EMBL/GenBank databases.  
 Yoshitaka Sano, Kyoto Institute of Technology, Department of  
 Applied Biology; Matsugasaki, Sakyo-ku, Kyoto, Kyoto 606, Japan  
 (E-mail:ysano@ipc.kit.ac.jp, Tel:075-724-7764, Fax:075-724-7764)  
 2 (sites)  
 Sano.Y., Wada.M., Hashimoto.Y., Matsumoto.T. and Kojima.M.  
 Sequences of ten circular ssDNA components associated with the milk  
 vetch dwarf virus genome  
 J Gen. Virol. 79 (Pt 12), 3111-3118 (1998)  
 JOURNAL 99094638  
 MEDLINE Location/Qualifiers  
 FEATURES  
 source  
 1..1022  
 /organism="milk vetch dwarf virus"  
 /specific\_host="Astragalus sinicus"  
 /db\_xref="taxon:67585"  
 /chromosome="segment 10"  
 /lab\_host="Pisum sativum"  
 /note="isolated from sieve element of broadbean"  
 1..29  
 stem\_loop  
 57..908  
 gene  
 /gene="Rep"  
 57..908  
 CDS  
 /gene="Rep"  
 /function="site-specific DNA nicking and joining"  
 /note="putative viral replication-initiation protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="replication initiation protein"  
 /protein\_id="BA034048.1"  
 /db\_xref="GI:3798738"  
 /translation="MPSIRAIHWCFTLNFSGKIPEIVWTADVQYSIQHVRVNHDLQ



|              |  |  |             |
|--------------|--|--|-------------|
| QY           | 552  | ttgaaatcaacagggcagaataaaactgtggaatatatttaaagtatgtcatcattaa | 611         |
| Df           | 876  | TCTAAATTCCTCGAAGATAGAATAAGATTATTATTGTTAAAAAGCAAAATTATATTG  | 935         |
| QY           | 612  | tacaccaataccqccgcg   | 630         |
| Df           | 936  | CACATGACAATCGTAGC  | 954         |
| <br>         |  |  |             |
| RESULT       | 12   |  |             |
| FBI32185     | 1015 bp  | DNA circular VPL   | 16-SEP-2000 |
| LOCUS        | faba bean necrotic yellows virus C7-Eg gene, isolate Egyptian  |  |             |
| DEFINITION   | EVI-93.  |  |             |
| ACCESSION    | AJ132185   |  |             |
| VERSION      | AJ132185.1   | GI:4995167   |             |
| KEYWORDS     | C7-Eg gene; rep protein.   |  |             |
| SOURCE       | faba bean necrotic yellows virus.  |  |             |
| ORGANISM     | faba bean necrotic yellows virus   |  |             |
| REFERENCE    | Viruses; ssDNA viruses; Nanovirus.   |  |             |
| AUTHORS      | 1 (bases 1 to 1015)  |  |             |
| TITLE        | Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.<br>Ten distinct circular ssDNA components, four of which encode<br>putative replication-associated proteins, are associated with<br>faba bean necrotic yellows virus genome  |  |             |
| JOURNAL      | J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)  |  |             |
| MEDLINE      | 99094637   |  |             |
| REFERENCE    | 2 (bases 1 to 1015)  |  |             |
| AUTHORS      | Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetten,H.J.<br>and Gronenborn,B.  |  |             |
| TITLE        | A single rep protein initiates replication of multiple genome<br>components of faba bean necrotic yellows virus, a single-strand<br>DNA virus of plants  |  |             |
| JOURNAL      | J. Virol. 73 (12), 10173-10182 (1999)  |  |             |
| MEDLINE      | 20027244   |  |             |
| REFERENCE    | 3 (bases 1 to 1015)  |  |             |
| AUTHORS      | Katul,L.   |  |             |
| TITLE        | Direct Submission  |  |             |
| JOURNAL      | Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie,<br>Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. L.  |  |             |
| MEDLINE      | u. Forstwirtschaft, Messeweg 11 -12, 38104 Braunschweig, GERMANY   |  |             |
| REFERENCE    | Location/Qualifiers  |  |             |
| FEATURES     | 1..1015  |  |             |
| source       | /organism="faba bean necrotic yellows virus"   |  |             |
|              | /virion  |  |             |
|              | /isolate="Egyptian EVI-93"   |  |             |
|              | /db_xref="taxon:59817"   |  |             |
|              | /lab_host="Vicia faba"   |  |             |
|              | 1..33  |  |             |
| stem_loop    | 60..1014   |  |             |
| gene         | J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)  |  |             |
| CDS          | 60..911  |  |             |
|              | /genes="C7-Eg"   |  |             |
|              | /function="replication-associated protein"   |  |             |
|              | /note="33.05 kDa"  |  |             |
|              | /codon_start=1   |  |             |
|              | /evidence=experimental   |  |             |
|              | /product="Rep protein"   |  |             |
|              | /protein_id="CAF44025.1"   |  |             |
|              | /db_xref="GI:4995168"  |  |             |
|              | /db_xref="SPTRMBL:Q9WIK0"  |  |             |
|              | /translation="MPSIRATHWCFTLNFSGSIPEINWTADVOYSIWHQERVGHH<br>GYIQMNKHVTLLKKMKLELPGLAHLEMAKAPKAIEYCQKKSAIAGPWEGTGWISS<br>RKLMFERFDGPEHKGLEDGPLYRCLRSVOMKK IRESCTWNFDLRPMODELTKT<br>EPDPTDI IIVYGPHGGECKGSFAKYLTLKEGWWYTAGKATDMLYSIDPCTCHV<br>IPRCTREEYINVSVIEQIKNRVININTKYECTIRDDGHNVHVFCNFPDPVTKRI<br>RIKLINC" |  |             |
| TATA_signal  | 979..984   |  |             |
|              | /gene="C7-Eg"  |  |             |
| polyA_signal | 1009..1014   |  |             |
|              | /gene="C7-Eg"  |  |             |
| BASE COUNT   | 346 a 168 c 228 g 273 t  |  |             |

\_\_\_\_\_



us-09-462-955-1\_copy\_383\_1215.rge

Thu Oct 25 13:08:01 2001

...

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:34 ; Search time 5701.85 Seconds  
(without alignments)  
1144.785 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_583\_1004  
Perfect score: 422  
Sequence: 1 cctgaagtgtgtctgcgat.....catctaaattacaccaatac 422

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_om.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_pl1.\*
- 13: gb\_pl2.\*
- 14: gb\_pl3.\*
- 15: gb\_pl4.\*
- 16: em\_ba1.\*
- 17: em\_ba2.\*
- 18: em\_fun.\*
- 19: em\_htgo\_hum.\*
- 20: em\_htgo\_inv.\*
- 21: em\_htgo\_rod.\*
- 22: em\_htg\_hum1.\*
- 23: em\_htg\_hum2.\*
- 24: em\_htg\_hum3.\*
- 25: em\_htg\_hum4.\*
- 26: em\_htg\_hum5.\*
- 27: em\_htg\_hum6.\*
- 28: em\_htg\_hum7.\*
- 29: em\_htg\_hum8.\*
- 30: em\_htg\_inv1.\*
- 31: em\_htg\_inv2.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_rod.\*
- 34: em\_hum1.\*
- 35: em\_hum2.\*
- 36: em\_hum3.\*
- 37: em\_hum4.\*
- 38: em\_hum5.\*
- 39: em\_hum6.\*
- 40: em\_hum7.\*
- 41: em\_in.\*
- 42: em\_om.\*
- 43: em\_or.\*

- 44: em\_ov.\*
- 45: em\_pat.\*
- 46: em\_ph.\*
- 47: em\_pl.\*
- 48: em\_ro.\*
- 49: em\_sts.\*
- 50: em\_sy.\*
- 51: em\_un.\*
- 52: em\_vi.\*
- 53: gb\_sts1.\*
- 54: gb\_sts2.\*
- 55: gb\_sts3.\*
- 56: gb\_sy.\*
- 57: gb\_un.\*
- 58: gb\_vi1.\*
- 59: gb\_vi2.\*
- 60: gb\_htg1.\*
- 61: gb\_htg2.\*
- 62: gb\_htg3.\*
- 63: gb\_htg4.\*
- 64: gb\_htg5.\*
- 65: gb\_htg6.\*
- 66: gb\_htg7.\*
- 67: gb\_htg8.\*
- 68: gb\_htg9.\*
- 69: gb\_htg10.\*
- 70: gb\_htg11.\*
- 71: gb\_htg12.\*
- 72: gb\_htg13.\*
- 73: gb\_htg14.\*
- 74: gb\_htg15.\*
- 75: gb\_htg16.\*
- 76: gb\_htg17.\*
- 77: gb\_htg18.\*
- 78: gb\_htg19.\*
- 79: gb\_htg20.\*
- 80: gb\_htg21.\*
- 81: gb\_htg22.\*
- 82: gb\_htg23.\*
- 83: gb\_htg24.\*
- 84: gb\_htg25.\*
- 85: gb\_pr1.\*
- 86: gb\_pr2.\*
- 87: gb\_pr3.\*
- 88: gb\_pr4.\*
- 89: gb\_pr5.\*
- 90: gb\_pr6.\*
- 91: gb\_pr7.\*
- 92: gb\_pr8.\*
- 93: gb\_pr9.\*
- 94: gb\_rol.\*
- 95: gb\_ro2.\*
- 96: gb\_in4.\*
- 97: gb\_pr10.\*
- 98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description        |
|------------|-------|---------|--------------|-------|--------------------|
| 1          | 422   | 100.0   | 1291         | 58    | CFDCG              |
| 2          | 117.6 | 27.9    | 1007         | 58    | FBEL32187          |
| 3          | 116   | 27.5    | 1004         | 59    | NYV5966            |
| 4          | 114.2 | 27.1    | 1096         | 9     | AR063452           |
| 5          | 114.2 | 27.1    | 1096         | 58    | BYTV2              |
| 6          | 112.6 | 26.7    | 1091         | 9     | AR063453           |
| 7          | 112.4 | 26.6    | 1095         | 58    | AF216222           |
| 8          | 103.2 | 24.5    | 1111         | 58    | BBU12586           |
|            |       |         |              |       | M29963 Coconut fol |
|            |       |         |              |       | AJ132187 faba bean |
|            |       |         |              |       | AJ005966 faba bean |
|            |       |         |              |       | AR063452 Sequence  |
|            |       |         |              |       | L32167 Banana bunc |
|            |       |         |              |       | AR063453 Sequence  |
|            |       |         |              |       | AF216222 Banana bu |
|            |       |         |              |       | U12586 Banana bunc |

Thu Oct 25 13:08:14 2001

TPGVKRLAQRFAEEDPDLRDPGGYRRVCHGASVEVTHAAENPFPPYHNQ  
LEVLGAIEPADDRITLWICRGDGDGKSVFAKYLGLKPDWFTYCGTRKDLVQYIE  
DKRNLILIDVPCNLEVLNALLECVKNRAFSDDKYEPLSYLGFHDVHVLVFAVFLPD  
YLKTSRDRILKWN1  
314..775  
/note="ORF 2"  
/codon\_start=1  
/protein\_id="AAA42895.1"  
/db\_xref="GI:323308"  
/translation="MTGETWRSRPVPTNRITVVRNGCFSSSTESRLVLESKGDWPN  
DLRLNMSANKTQADTALYELRWNGQDPLKIRSHPHITIGSLKCCLESQRT  
IAQSGYADETEETGSPCLPNISDSSPTGSHVVEPERTYCTSTRTQNEI"  
complement(422..568)  
/note="ORF 6"  
/codon\_start=1  
/protein\_id="AAA42896.1"  
/db\_xref="GI:323309"  
/translation="MEMGTDFORPILSPPLKRVQIRFIRGLGLPGGVHQPQIVGP  
IVAF"  
639..797  
/note="ORF 3"  
/codon\_start=1  
/protein\_id="AAA42897.1"  
/db\_xref="GI:323310"  
/translation="METRRRRREVRCQISRTQARLVLMHWNQKGRIVPVHRGPKTK  
FNPRCTQV"  
complement(823..987)  
/note="ORF 5"  
/codon\_start=1  
/protein\_id="AAA42898.1"  
/db\_xref="GI:323311"  
/translation="MTHLNIPQFVSVPADFOIIRIDGKYEYMHMVPEKITKGFFV  
RTCPVLNTP"  
1098..1286  
/note="ORF 4"  
/codon\_start=1  
/protein\_id="AAA42899.1"  
/db\_xref="GI:323312"  
/translation="MNRVMGGPTIKDSIWRNMLLCLCLOCTQPLSTSPQVSSLLLEKK  
AASLYLPSICFCAIGRLS"  
BASE COUNT 336 a 323 c 332 g 300 t  
ORIGIN

U12587 Banana bunc  
AJ132185 faba bean  
AF216221 Banana bu  
AJ005964 faba bean  
AR063451 Sequence  
L32166 Banana bunc  
U16735 Subterranea  
X80879 FBMY gene  
AB000922 Milk vetc  
AB000921 Milk vetc  
AB000947 Milk vetc  
U02312 Banana bunc  
AJ005968 faba bean  
AB000920 Milk vetc  
U16731 Subterranea  
AJ132345 Nanovirus  
AJ238493 Nanovirus  
AJ132344 Nanovirus  
AR102330 Sequence  
AR112007 Sequence  
AR010227 Sequence  
U18077 Banana bunc  
AR102225 Sequence  
AR010228 Sequence  
AR010240 Sequence  
AF102780 Banana bu  
S56276 orf V1 {com  
AR010232 Sequence  
AR010234 Sequence  
AR010229 Sequence  
AR010231 Sequence  
AF238874 Banana bu  
U97525 Banana bunc  
AR010233 Sequence  
AR010226 Sequence  
AF238875 Banana bu  
AF246123 Banana bu

## ALIGNMENTS

RESULT 1  
CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993  
LOCUS Coconut foliar decay virus, complete genome.  
ACCESSION M29963  
VERSION M29963.1 GI:323306  
KEYWORDS circular; complete genome.  
SOURCE Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.  
ORGANISM Coconut foliar decay virus  
VIRUSES: ssDNA viruses; Nanovirus.  
REFERENCE 1 (bases 1 to 1291)  
Rohde, W., Randles, J.W., Langridge, P. and Hanold, D.  
Nucleotide sequence of a circular single-stranded DNA associated  
with coconut foliar decay virus  
Virology 176, 648-651 (1990)  
MEDLINE 90266484  
COMMENT Draft entry and printed sequence for [1] kindly submitted by  
W. Rohde, 15-MAR-1989, for release after publication.  
FEATURES  
source  
1..1291  
/organism="Coconut foliar decay virus"  
/db\_xref="taxon:12474"  
40..70  
/note="stem-loop structure"  
103..975  
/note="ORF 1"  
/codon\_start=1  
/protein\_id="AAA42894.1"  
/db\_xref="GI:323307"  
/translation="MGSSIRRWCFLLNVEETEAEAAVVRRIEISLNLVYVIGDEVAPS  
TQGRHLQGFILHTGRRLQGLKTVLGNDRHLEPTRGSDQNRDCKSERVLLHGVF

stem\_loop

CDS

```

|||||
Db 943 AGCAGGACAGAAATAAACTGTGGAATATTAAAGTATGTGTCACTAAATTACACCAAT 1002
QY 421 ac 422
Db 1003 AC 1004

RESULT 2
FBE132187
LOCUS FBE132187 1007 bp DNA circular VRL 16-SEP-2000
DEFINITION faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian
EVI-93.
ACCESSION AJ132187
VERSION AJ132187.1 GI:4995171
KEYWORDS C9-Eg gene; rep protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus.
REFERENCE 1 (bases 1 to 1007)
AUTHORS Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
99094637
JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
MEDLINE
REFERENCE 1 (bases 1 to 1007)
AUTHORS Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetten,H.J. and Gronenborn,B.
TITLE A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
J. Virol. 73 (12), 10173-10182 (1999)
20027244
JOURNAL J. Virol. 73 (12), 10173-10182 (1999)
MEDLINE
REFERENCE 3 (bases 1 to 1007)
AUTHORS Katul,L.
TITLE Direct Submission
Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. Land- u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
FEATURES
source
1..1007
/organism="faba bean necrotic yellows virus"
/virion
/isolate="Egyptian EVI-93"
/db_xref="taxon:59817"
/lab_host="Vicia faba"
1..39
stem_loop
gene
73..1003
/gene="C9-Eg"
73..918
/gene="C9-Eg"
/function="replication-associated protein"
/notes="32.53 kba"
/codon_start=1
/evidence-experimental
/product="Rep protein"
/protein_id="CAB44027.1"
/db_xref="GI:4995172"
/db_xref="SPTREMBL:Q9WIK2"
/translation="MSAVNVVFTLNFAGEVPLSFDERVOYAVWQHERVNHDTQGV
QLKKAKMTVKNIIGGNPHLEKMKGSIEASAYAKDESRVGPWSYVELLKGSHK
RKIMELIKDPNELEPOKYRRAMAWSAMDESRLAEGGPPYTLISWQETVILGLLEE
EPNDRILITWYGPNGECKSOPGKFLGLKLDYLLPGKTDQDMYMLMKPNKANVYND
IPRCNSEYLVQFMELIKNRTIPSKYKPEVPGCIINNKIHVILVANLVPDEKISQDRI
KTIYC"
TATA_signal 939..944
polyA_signal 998..1003
BASE COUNT 353 a 161 c 227 g 266 t
ORIGIN

```

```

Query Match 27.9%; Score 117.6; DB 58; Length 1007;
Best Local Similarity 56.4%; Pred. No. 2.8e-23;
Matches 219; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 6 agtgctgtcgtcgatcgagaccagcgaggaatcgcacaaatcctctgatatcggaag 65
|||||
Db 531 AGTGTGGGCTATTAGAGAGAGAGCCCAATGACCGCTATTATTGGGTCTACGGGCC 590
|||||
QY 66 agacgagggagagcggaagtcggtttgccaaatatctcggactcaagcccgactggt 125
|||||
Db 591 AAATGTTAAGCAAGAGAAATACACAGTTTGTGTAATTCCTGGGATTAATAAAGATTACCT 650
|||||
QY 126 ctacacatgtgtggaaccagaaagacattattgtaccagtcacatcgaggagcccaaacg 185
|||||
Db 651 TTATTTACCTGGAGGTAACCAACCAAGATATGACATATATGTAAATGAAAAATCCAAAGGC 710
|||||
QY 186 aaatttaactcgtatgtaccaggtgttaatttagagatttaaaattatgcctgttaga 245
|||||
Db 711 AAATGTTGATGGATATTCCTCGTTGTAATTCCTGAATTTAAATTTATCAATTTATGGA 770
|||||
QY 246 atgtgttaagacagggcattcagtcggagacaaatacaccoccttagttatcttggtt 305
|||||
Db 771 ATTAATTAATAAGAACCATTTAGTTATAAATATGAACGAGTTGGATGATTATATAA 830
|||||
QY 306 cgaccatgtcgtactcgtatttgcgaatgctcgtcgtatatttgaataatcagcag 365
|||||
Db 831 TAATAAAATACATGTAATTTAGTTAGCTAATGTTATGCTGCTGATTGAAAAATATTGCA 890
|||||
QY 366 gacagaataaaactgtggaatatttaa 393
|||||
Db 891 GCACAGAAATTAATAATTTATTGTGTA 918
|||||

RESULT 3
NTV5966
LOCUS NTV5966 1004 bp DNA circular VRL 13-JAN-1999
DEFINITION faba bean necrotic yellows virus C9 gene.
ACCESSION AJ005966
VERSION AJ005966.1 GI:3550532
KEYWORDS C9 gene; component 9; putative; rep protein; replication associated protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus.
REFERENCE 1 (bases 1 to 1004)
AUTHORS Katul,L.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
REFERENCE 2 (bases 1 to 1004)
AUTHORS Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
99094637
FEATURES
Location/Qualifiers
1..1004
/organism="faba bean necrotic yellows virus"
/virion
/isolate="SV292-88"
/db_xref="taxon:59817"
73..918
/gene="C9"
73..918
/gene="C9"
/function="putative replication associated (rep) protein"
/product="component 9"
/protein_id="CAA06789.1"
/db_xref="GI:3550533"

```

```

/db_xref="SPTREMBL:O91252"
/translation="MSAVNPFVTLNFACEVPLSEDFVQVAVQHERVNDHDTQGVY
QLKKAKAMTNTKNIIGNPHLMKNGSEESASATAQKEESRVAGPWSYGLLKKGSHK
RKTMLTIDPWESESPQRYRRAMAWSAMDESRKLAEEEGPFTMYFSWQETVGLLEL
EMENDTIIWYFNGNEGKSGFKGLKDYLYLPGGKTODMYTLMKMKPFKANVYMD
IPRCNSLYNQEMELIKNRTIYSKYEPVGCIIINNKHIVILANVLDPDEKTSQDRI
KIYYC"
BASE COUNT 352 a 159 c 226 g 267 t
ORIGIN
Query Match 27.5%; Score 116; DB 59; Length 1004;
Best Local Similarity 56.2%; Pred. No. 8.1e-23;
Matches 218; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 6 agtgctgtctgcgtcgagagcgagcgagcgatcgacaaatcctctgatatcgcgacg 65
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 AGTGTGGGCCCTATTAGAAGAGAGGCCCAATGACCGTACTATTATTTGGCTACGGCCC 590
QY 66 agcggagagcgaggaagtcggttttggccaaatatctcggaactaaagcccgactggtt 125
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 AATGGTAAATGAAGAAATACACAGTTTGGTAAATTCCTGGGATTAATAAAGATTACCT 650
QY 126 ctacacatgtgtggaaccagaagaagcgtattctaccagtacatcgagaccacaaacg 185
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 TTATTACCTGAGAGTAAACCCCAAGATATGACATATATGTGTAATGAAATAATCCAAAGGC 710
QY 186 aaatttaactctcgatgccaccagtgtaatttagagatttaaatattgccctgttaga 245
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 AATGTGTGATGATATTCCTCGTTGTAATCTCAATATTAAATACCAATTTATGGA 770
QY 246 atgtgttaagaacagggcattcagttcggaacaaatcacaccctctagttatcttgggtt 305
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 ATTAATTTAAATAATGAACCATATATAGTTTAAATATGAACCACTGGATGATATATAA 830
QY 306 cgaccatgtgcattgctgatttggccaaatgctgcctgatttattgaaatcagcag 365
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 TAATAAATACATGTAATTGTATTAGCTAATGTATTGCTGATTATGAAAAAATTAGTCA 890
QY 366 ggacagataaataaactgtggaaatttaa 393
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 891 GGATAGAAATTAATAATTTATTGTTAA 918
. . .

RESULT 4
AR063452 LOCUS AR063452 1096 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 4 from patent US 5846705.
ACCESSION AR063452.1 GI:5992760
VERSION
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1096)
AUTHORS Wu, R., You, L. and Soong, T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus
JOURNAL Patent: US 5846705-A 08-DEC-1998;
FEATURES Location/Qualifiers
source 1. 1096
/organism="unknown"
BASE COUNT 347 a 231 c 244 g 274 t
ORIGIN
Query Match 27.1%; Score 114.2; DB 9; Length 1096;
Best Local Similarity 57.5%; Pred. No. 2.7e-22;
Matches 225; Conservative 0; Mismatches 163; Indels 3; Gaps 0;

QY 19 atcggagagccagcgagatcgacaaatcctctgatatcgcgagcgagcgagagac 78
AR063452 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



| ORIGIN                    |     | Query Match  |     | 26.7%; Score 112.6; DB 9; Length 1091; |  |
|---------------------------|-----|--|-----|--|--|
| Best Local Similarity     |     | 57.3%; Pred. No. 7.8e-22;                                      |     |  |  |
| Matches 224; Conservative |     | 0; Mismatches 164; Indels                                      |     | 3; Gaps                                |  |
| Qy                        | 19  | atcggagagccagcgagcgtatcgcaaatctcttggtatcgagcagcagcgagcgagac    | 78  |  |  |
| Db                        | 546 | ATCGAAGGTGTTCTGATGATCGAAGATATCATCTGGGTATACGCTCTCCGGAGCGCAA     | 605 |  |  |
| Qy                        | 79  | gggaagtcctgtgttgcgaatatctcggactcaagcccgactgtgtctacacatgtggt    | 138 |  |  |
| Db                        | 606 | GCAAAGTCAACCTTCGCAAGATATCTATCATATAAACCTGGATGGGATATATCAACGCT    | 665 |  |  |
| Qy                        | 139 | ggaaccaggaaggcgtattgtaccagtcacgtacatcgagcagcccaaacgaaattaatctc | 198 |  |  |
| Db                        | 666 | GGAAAGCGTCGGATATGATCCACATCATACGATGGATCCTGATAATCATTTGGATTATT    | 725 |  |  |
| Qy                        | 199 | gatgtaccocagggtgaatttagagtgatttaaatatgccctgttagaattgtttaagaac  | 258 |  |  |
| Db                        | 726 | GATATCCCAAGAGTCATCTCAGATTAATCTGATTAATGATTAATTTTCAAGATAGAATA    | 905 |  |  |
| Qy                        | 259 | aggcattcagtcgacgaataacgaacc--ccttagttatcttgggttcgcacctgtg      | 315 |  |  |
| Db                        | 786 | AGAGTTTATAATAATCAAAATACGAACCATGTGTATTAGAAAAAGATGCACAAAATGTC    | 845 |  |  |
| Qy                        | 316 | catgtactgtatttgccaatgctcgtcgctgattatttgaataatcagcagggacagaata  | 375 |  |  |
| Db                        | 846 | CATGTAATTTGTTATGCAAAATGTTGCCGTGATTATGTAATAATTTTCAAGATAGAATA    | 905 |  |  |
| Qy                        | 376 | aaactgtggaattttaaagtatgtgtcatc                                 | 406 |  |  |
| Db                        | 906 | AAATAATTAATTGTTGAGAAAGGAAACTTC                                 | 936 |  |  |

| RESULT 7   |  | AF216222   |  | 1095 bp DNA |  | VRL |  | 08-MAR-2001 |  |
|------------|--|--|--|-------------|--|-----|--|-------------|--|
| LOCUS      |  | Banana bunchy top virus satellite S2 replication initiation protein  |  |             |  |     |  |             |  |
| DEFINITION |  | (ORF V1) gene, complete cds.   |  |             |  |     |  |             |  |
| ACCESSION  |  | AF216222   |  |             |  |     |  |             |  |
| VERSION    |  | AF216222.1 GI:12004327   |  |             |  |     |  |             |  |
| KEYWORDS   |  | Banana bunchy top virus.   |  |             |  |     |  |             |  |
| SOURCE     |  | Banana bunchy top virus.   |  |             |  |     |  |             |  |
| ORGANISM   |  | Viruses; ssDNA viruses; Circoviridae.  |  |             |  |     |  |             |  |
| REFERENCE  |  | 1 (bases 1 to 1095)  |  |             |  |     |  |             |  |
| AUTHORS    |  | Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.  |  |             |  |     |  |             |  |
| TITLE      |  | Additional rep-encoding DNAs associated with Banana bunchy top virus   |  |             |  |     |  |             |  |
| JOURNAL    |  | Arch. Virol. 146 (1), 71-86 (2001)   |  |             |  |     |  |             |  |
| MEDLINE    |  | 21163975   |  |             |  |     |  |             |  |
| REFERENCE  |  | 2 (bases 1 to 1095)  |  |             |  |     |  |             |  |
| AUTHORS    |  | Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.  |  |             |  |     |  |             |  |
| TITLE      |  | Direct Submission  |  |             |  |     |  |             |  |
| JOURNAL    |  | Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia  |  |             |  |     |  |             |  |
| FEATURES   |  | location/Qualifiers<br>1..1095<br>/organism="Banana bunchy top virus"<br>/isolate="Taiwan"<br>/db_xref="taxon:12585"<br>/note="satellite S2"<br>64..921<br>/gene="ORF V1"<br>64..921<br>/gene="ORF V1"<br>64..921<br>/gene="ORF V1"<br>/codon_start=1<br>/product="replication initiation protein" |  |             |  |     |  |             |  |
| mRNA       |  |  |  |             |  |     |  |             |  |
| gene       |  |  |  |             |  |     |  |             |  |
| CDS        |  |  |  |             |  |     |  |             |  |



[illegible]

QKHLQGLSLKKSFKLGLKKYSRAHWERARGSDENAKYSKETTLLLE  
GNNRKLSEMVSRSPERMRIEQPIETHRYTSVKLKKFKEFVHPCLDHPWQ  
IDEEPDRSIIWYGPNGNEGKSTYAKSLMKKDWFYTRGGKKENILFSYVDE

|      |            |
|------|------------|
| gene | 00.1.2010  |
|      | /gene="C7" |
| CDS  | 60.911     |



Viruses: ssDNA viruses; Nanovirus.

1 (bases 1 to 1017)  
 Boevink, P., Chu, P.W. and Keese, P.  
 Sequence of subterranean clover stunt virus DNA: affinities with  
 the geminiviruses  
 Virology 207 (2), 354-361 (1995)  
 5193233  
 2 (bases 1 to 1017)  
 Boevink, P.C.  
 Direct Submission  
 Submitted (28-Oct-1994) Petra C. Boevink, C.S.I.R.O. Division of  
 Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia  
 Location/Qualifiers  
 1..1017  
 /organism="Subterranean clover stunt virus"  
 /isolate="F"  
 /specific\_host="Trifolium subterraneum L."  
 /db\_xref="taxon:36772"  
 /chromosome="component 6"  
 /lab\_host="Trifolium subterraneum cv. Mt. Barker"  
 48..905  
 /gene="SCSV6"  
 48..905  
 /gene="SCSV6"  
 /note="possible replication associated protein"  
 /codon\_start=1  
 /protein\_id="AA68022.1"  
 /db\_xref="GI:571436"  
 /translation="MPTQSRWSWPTLNFEGEILIPFNESVOYACQHERVGHDLHQ  
 GSHKRLRFRFEDEEMKIDPISLYRRLCSRKMAEQKSSSEWNTDLRPQDEVMVH  
 LEPEPYRTILWVYGPAGNEGKSTFARHLSLKDGWGLPGKTKQDMHLVTAEPKNN  
 VFDPVRSSEYNYGVIEQVKRNVN"KYEPVMDNDHPVHVIVFANVLPDLGLKS  
 EDRIKLIRNC"  
 stem\_loop  
 join(1005..1017,1..19)  
 /note="region of sequence and predicted structural  
 similarity with the conserved stem-loop sequence in the  
 intergenic region of the geminiviruses, which is the  
 origin of replication"  
 BASE COUNT 312 a 160 c 265 g 280 t  
 ORIGIN  
 Query Match 22.6%; Score 95.4; DB 59; Length 1017;  
 Best Local Similarity 54.1%; Pred. No. 7.1e-17;  
 Matches 217; Conservative 0; Mismatches 181; Indels 3; Gaps 1;  
 QY 4 gaagtgtctgtcgatcgaggagccaggcagcagatcgacaaatcctcggatcgga 63  
 Db 513 GAAGTCATGCATTTGTTAGAGGAGAACACAGATATAGAACGATAATCTGGGTG 572  
 QY 64 cgagcagaggagcgggaagtcggtgttgccaaatctcggactcaagcccgactgg 123  
 Db 573 CTTGCTGTATGAGGCAATCTACATTTGCAAGACATCTGTCATTGAAGATGTTGG 632  
 QY 124 ttctacacatgtggttggaacccagaaaggaagcgtatgtaccagtcacatcgagacc 183  
 Db 633 GGTATCTCCCTGGAGGAGAACACAGATATGATGTCATCTTGACTGCTGAGCCTA 692  
 QY 184 cgaaatttaactcgtatgaccaggtgaatttagagtattaaattatcgccctgttta 243  
 Db 693 AATAAATGGGTATTTCACATACCCAGTAGTACAGTATGTAATATGTTGGTAATA 752  
 QY 244 gaatgtttaaagacaggcgcattcagttcggaacaaatcacgaacc---cctagttatc 300  
 Db 753 GAACAGTTAAGAATAGGGTAAATGTTGAATACTAAGTATGAGCCATGTGTAATG 812  
 QY 301 gggtcgaccatgtgcattactcgtatttggcaatgctcctgctgattttgaaatc 360  
 Db 813 GATAAATCCTCTGTTTCATGTAATGTTGTTGCAAAAGTACTCCAGATTTGGGAA 872  
 QY 361 agcaggagacagaataaaactgtggaattttaaagtatgtg 401

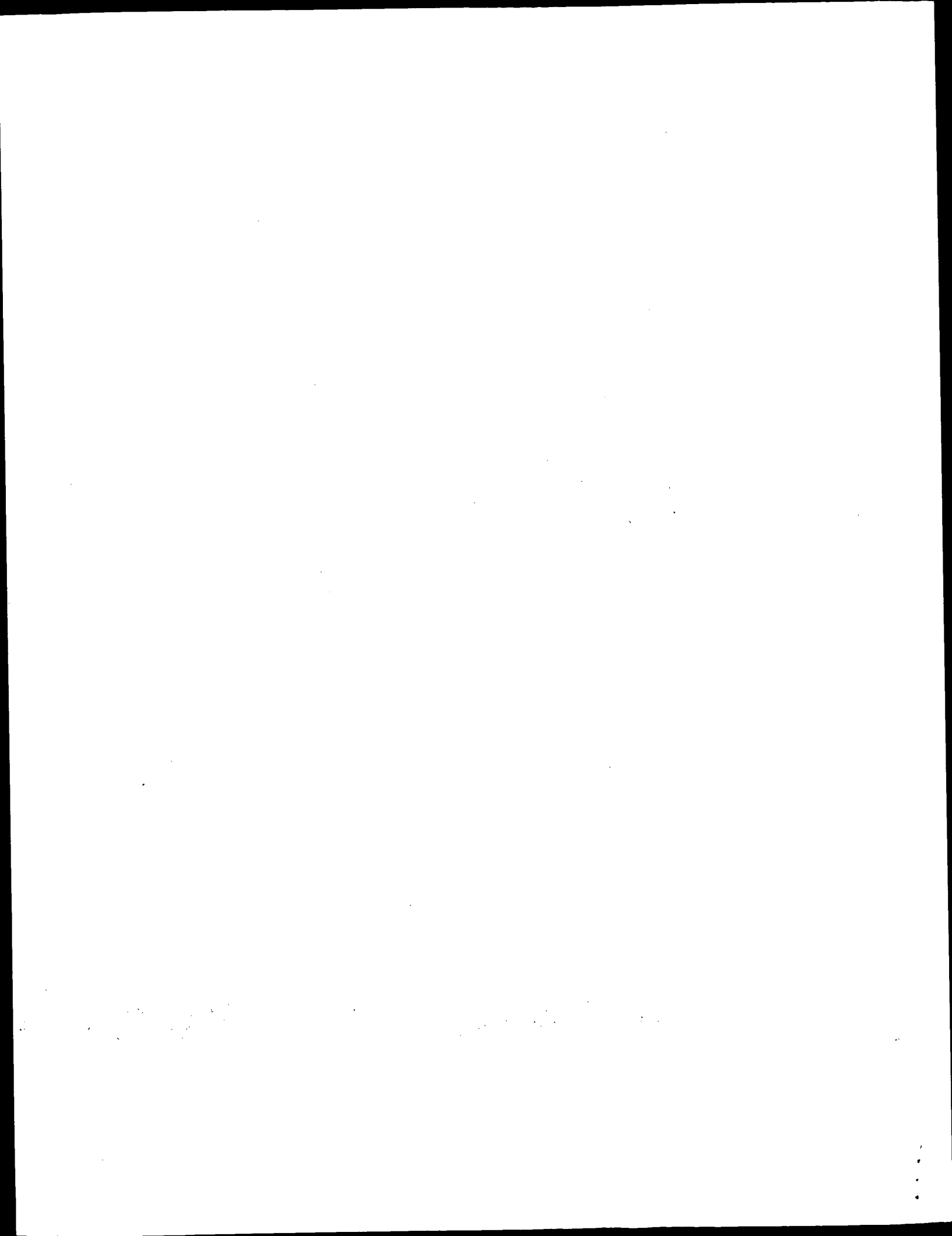
Thu Oct 25 13:08:14 2001

us-09-462-955-1\_copy\_583\_1004.rge

Page 11

Db 873 AGTGAAGATAGATAAAATTAATTCGTTGTTGAAACTCTG 913

Search completed: October 24, 2001, 11:42:35  
Job time: 6340 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:08 ; Search time 5701.85 Seconds  
(without alignments)  
84.096 Million cell updates/sec

Title: US-09-462-955-1-copy\_941\_971  
Perfect score: 31  
Sequence: 1 tcagcaggacagataaaactgtggaatat 31

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_ba3:\*
- 4: gb\_in1:\*
- 5: gb\_in2:\*
- 6: gb\_in3:\*
- 7: gb\_om:\*
- 8: gb\_ov:\*
- 9: gb\_pat1:\*
- 10: gb\_pat2:\*
- 11: gb\_ph:\*
- 12: gb\_pl1:\*
- 13: gb\_pl2:\*
- 14: gb\_pl3:\*
- 15: gb\_pl4:\*
- 16: em\_ba1:\*
- 17: em\_ba2:\*
- 18: em\_fun:\*
- 19: em\_htgo\_hum:\*
- 20: em\_htgo\_inv:\*
- 21: em\_htgo\_rod:\*
- 22: em\_htg\_hum1:\*
- 23: em\_htg\_hum2:\*
- 24: em\_htg\_hum3:\*
- 25: em\_htg\_hum4:\*
- 26: em\_htg\_hum5:\*
- 27: em\_htg\_hum6:\*
- 28: em\_htg\_hum7:\*
- 29: em\_htg\_hum8:\*
- 30: em\_htg\_inv1:\*
- 31: em\_htg\_inv2:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_rod:\*
- 34: em\_hum1:\*
- 35: em\_hum2:\*
- 36: em\_hum3:\*
- 37: em\_hum4:\*
- 38: em\_hum5:\*
- 39: em\_hum6:\*
- 40: em\_hum7:\*
- 41: em\_in:\*
- 42: em\_om:\*
- 43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_vil2:\*

60: gb\_vil3:\*

61: gb\_vil4:\*

62: gb\_vil5:\*

63: gb\_vil6:\*

64: gb\_vil7:\*

65: gb\_vil8:\*

66: gb\_vil9:\*

67: gb\_vil10:\*

68: gb\_vil11:\*

69: gb\_vil12:\*

70: gb\_vil13:\*

71: gb\_vil14:\*

72: gb\_vil15:\*

73: gb\_vil16:\*

74: gb\_vil17:\*

75: gb\_vil18:\*

76: gb\_vil19:\*

77: gb\_vil20:\*

78: gb\_vil21:\*

79: gb\_vil22:\*

80: gb\_vil23:\*

81: gb\_vil24:\*

82: gb\_vil25:\*

83: gb\_vil26:\*

84: gb\_vil27:\*

85: gb\_vil28:\*

86: gb\_vil29:\*

87: gb\_vil30:\*

88: gb\_vil31:\*

89: gb\_vil32:\*

90: gb\_vil33:\*

91: gb\_vil34:\*

92: gb\_vil35:\*

93: gb\_vil36:\*

94: gb\_vil37:\*

95: gb\_vil38:\*

96: gb\_vil39:\*

97: gb\_vil40:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 31    | 100.0       | 1291   | 58    | CFDCG              |
| 2          | 21.2  | 68.4        | 137949 | 88    | AC022137           |
| c          | 21.2  | 68.4        | 162107 | 83    | AP002423           |
| 4          | 21.2  | 68.4        | 163192 | 76    | AC079751           |
| c          | 21.2  | 68.4        | 163192 | 76    | AC079751           |
| 6          | 21.2  | 68.4        | 177748 | 64    | AC016185           |
| c          | 21.2  | 68.4        | 181041 | 77    | AC087677           |
| 8          | 21.2  | 68.4        | 181239 | 68    | AC023983           |
|            |       |             |        |       | M29963 Coconut fol |
|            |       |             |        |       | AC022137 Homo sapi |
|            |       |             |        |       | AP002423 Homo sapi |
|            |       |             |        |       | AC079751 Homo sapi |
|            |       |             |        |       | AC079751 Homo sapi |
|            |       |             |        |       | AC016185 Homo sapi |
|            |       |             |        |       | AC087677 Homo sapi |
|            |       |             |        |       | AC023983 Homo sapi |

9 21 67.7 73465 92 HS256M13  
c 10 21 67.7 152318 79 AL162851  
c 11 21 67.7 168075 82 HS214G14  
c 12 21 67.7 340000 94 HS21C004  
c 13 20.8 67.1 214348 66 AC021220  
c 14 20.6 66.5 1211 94 AB02952S11  
c 15 20.6 66.5 4621 10 E13044  
c 16 20.6 66.5 4621 10 D21655  
c 17 20.6 66.5 4621 94 E78270  
c 18 20.6 66.5 178010 77 AC087067  
c 19 20.4 65.8 866 53 CNS07508  
c 20 20.4 65.8 101856 70 AC026725  
c 21 20.4 65.8 110620 88 AC079045  
c 22 20.4 65.8 111489 70 AC026706  
c 23 20.4 65.8 113853 89 AL137853  
c 24 20.4 65.8 124474 90 AL158012  
c 25 20.4 65.8 141514 80 AL359095  
c 26 20.4 65.8 153092 88 AC069475  
c 27 20.4 65.8 157069 67 AC022839  
c 28 20.4 65.8 159774 79 AL160173  
c 29 20.4 65.8 161545 81 AL451047  
c 30 20.4 65.8 164782 81 AL513175  
c 31 20.4 65.8 167060 78 AL138778  
c 32 20.4 65.8 167314 72 AC055798  
c 33 20.4 65.8 167333 79 AL353714  
c 34 20.4 65.8 167685 70 AC027429  
c 35 20.4 65.8 171829 62 AC011454  
c 36 20.4 65.8 173077 68 AC024511  
c 37 20.4 65.8 173409 73 AC074173  
c 38 20.4 65.8 180129 75 AC074144  
c 39 20.4 65.8 184800 75 AC074144  
c 40 20.4 65.8 189876 90 AL357892  
c 41 20.4 65.8 200930 81 AL391831  
c 42 20.4 65.8 245465 79 AL157367  
c 43 20.4 65.8 245465 79 AL157367  
c 44 20.4 65.8 279647 91 CNS07EEW  
c 45 20.4 65.8 347503 89 AF279660

ALIGNMENTS

RESULT 1  
CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993  
LOCUS Coconut foliar decay virus, complete genome.  
DEFINITION M29963  
ACCESSION M29963  
VERSION 1 GI:323306  
KEYWORDS circular; complete genome.  
SOURCE Coconut foliar decay virus  
ORGANISM Coconut foliar decay virus  
Viruses; ssDNA viruses; Nanovirus.  
REFERENCE 1 (bases 1 to 1291)  
AUTHORS Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.  
TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus  
JOURNAL Virology 176, 648-651 (1990)  
MEDLINE 90266484  
COMMENT Draft entry and printed sequence for [1] kindly submitted by W.Rhode, 15-MAR-1989, for release after publication.  
FEATURES  
source 1. .1291  
/organism="Coconut foliar decay virus"  
/db\_xref="taxon:12474"  
stem\_loop 40..70  
CDS 103..975  
/note="stem-loop structure"  
/notes="ORF 1"  
/codon\_start=1  
/protein\_id="AAA42894.1"  
/db\_xref="GI:323307"  
/translation="MGSSIRRWCFITLNYETEAEANVVRRIESNLVYVIGDEVAPS  
TGORHLQGFHLTKTRRLQGLKTVIGNDRIHLEPTGSDQNRDYSKERVLLHGVGP

CDS  
314..775  
/note="ORF 2"  
/codon\_start=1  
/protein\_id="AAA42895.1"  
/db\_xref="GI:323308"  
/translation="MTGFTWSPVVTNRITVRRNGCFSTESRLVLESKGDWPN  
DLRLNMSAWKTQADTEDALTELRWQDQGLKIRSHFHTITGSLKCCCLRSORT  
IAOSSYADETETGSPCLPNTSDSPSTGSHVVEPERTYCTISRTQNEI"  
/note="ORF 6"  
/codon\_start=1  
/protein\_id="AAA42896.1"  
/db\_xref="GI:323309"  
/translation="MEMGTFDFORPILSPPKLRVORIFGIRLGLPGGVHQPQILVGP  
IVAF"  
639..797  
/note="ORF 3"  
/codon\_start=1  
/protein\_id="AAA42897.1"  
/db\_xref="GI:323310"  
/translation="MTRRRRRRVRCQISRTQARLVLMWNQKGRIVPVHRCPTK  
FNPRCTQV"  
complement(823..987)  
/note="ORF 5"  
/codon\_start=1  
/protein\_id="AAA42898.1"  
/db\_xref="GI:323311"  
/translation="MTHTLNIPQFYSVPADFQIIRODICKYEMHMHVEPKITKGFVVF  
RTECPVINTP"  
1098..1286  
/note="ORF 4"  
/codon\_start=1  
/protein\_id="AAA42899.1"  
/db\_xref="GI:323312"  
/translation="MNRVMGGPTIKDSIWIRTNLLCLOCTOPLSTPSIQVSSLLEKK  
AASLYLPSICFCAIGRLS"  
BASE COUNT 336 a 323 c 332 g 300 t  
ORIGIN

Query Match 100.0%; Score 31; DB 58; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 0.0028;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagcagggacagataaaactgtggaatat 31  
|||||  
Db 941 TCAGCAGGACAGATAAACTGTGGAATAT 971

RESULT 2  
AC022137 137949 bp DNA PRI 26-NOV-2000  
LOCUS Homo sapiens chromosome 19 clone CTD-2224J9, complete sequence.  
DEFINITION AC022137  
ACCESSION AC022137  
VERSION AC022137.5 GI:11079415  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 137949)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 137949)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

REFERENCE
AUTHORS   3 (bases 1 to 137949)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
Submitted (03-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS   4 (bases 1 to 137949)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT    On Nov 3, 2000 this sequence version replaced gi:9964763.
Draft sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >40 99.9% of Sequence;
Estimated Total Number of Errors is 0.4.
STS Content:
SHGC-84347 G60949
SHGC-103885 G57910.

FEATURES             source
   source
1..137949
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2224J9"
BASE COUNT    39207 a 31405 c 30153 g 37184 t
ORIGIN
Query Match      68.4%; Score 21.2; DB 88; Length 137949;
Best Local Similarity 88.5%; Pred. No. 74;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcacgaggacaaataaaactgtg 26
||||| ||||| ||||| ||||| |||||
Db 16523 TCAGGAGGTACAGAAATAAACAGTGG 16548

RESULT 3
AP002423/c
LOCUS      Homo sapiens chromosome 18 clone RP11-863L8 map 18q11.2, WORKING
DEFINITION DRAFT SEQUENCE, 42 unordered pieces.
ACCESSION  AP002423
VERSION     AP002423.2 GI:8439535
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-863L8.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 162107)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens 162,107 genomic DNA of 18q11.2
JOURNAL     Published Only in Database (2000) In press
REFERENCE   2 (bases 1 to 162107)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp,
Fax:81-42-778-9924)
COMMENT     On Jun 9, 2000 this sequence version replaced gi:8131687.
-----Genome Center
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information

```

```

Center project name: HumDraft18
Center clone name: RP11-863L8
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 143757 bases at least Q40
Consensus quality: 152091 bases at least Q30
Consensus quality: 155858 bases at least Q20
Insert size: 158007; sum-of-contigs
Quality coverage: 4.32x in Q20 bases; sum-of-contigs
-----

```

NOTE: This is a 'working draft' sequence. It currently consists of 42 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 12846 contig of 12846 bp in length
12947 23222 contig of 10276 bp in length
23323 31114 contig of 7792 bp in length
31215 38997 contig of 7783 bp in length
39098 46783 contig of 7686 bp in length
46884 54555 contig of 7772 bp in length
54756 59749 contig of 4994 bp in length
59850 65160 contig of 5311 bp in length
65261 71291 contig of 6031 bp in length
71392 75257 contig of 3866 bp in length
75358 79204 contig of 3847 bp in length
79305 83972 contig of 4668 bp in length
84073 88891 contig of 4819 bp in length
88992 93736 contig of 4745 bp in length
93837 98281 contig of 4445 bp in length
98382 102273 contig of 3892 bp in length
102374 106147 contig of 3774 bp in length
106248 110389 contig of 4142 bp in length
110490 112888 contig of 2399 bp in length
112989 115657 contig of 2669 bp in length
115758 119437 contig of 3680 bp in length
119538 123011 contig of 3474 bp in length
123112 125411 contig of 2300 bp in length
125512 128514 contig of 3003 bp in length
128615 131919 contig of 3305 bp in length
132020 133329 contig of 1310 bp in length
133430 136197 contig of 2768 bp in length
136298 138656 contig of 2359 bp in length
138757 140802 contig of 2046 bp in length
140903 142951 contig of 2049 bp in length
143052 145259 contig of 2208 bp in length
143360 147234 contig of 1875 bp in length
147335 149339 contig of 2005 bp in length
149440 150983 contig of 1544 bp in length
151084 152215 contig of 1132 bp in length
152316 153476 contig of 1161 bp in length
153577 155210 contig of 1634 bp in length
153711 156633 contig of 1323 bp in length
156734 157980 contig of 1247 bp in length
158081 159424 contig of 1344 bp in length
159525 160646 contig of 1122 bp in length
160747 162107 contig of 1361 bp in length

```

Sequence updated (06-Jun-2000).  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 42 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 12846: contig of 12846 bp in length
* 12847 12946: gap of 100 bp
* 12947 23222: contig of 10276 bp in length

```

Thu Oct 25 13:08:32 2001

\* 23223 23322: gap of 100 bp  
\* 23323 31114: contig of 7792 bp in length  
\* 31115 31214: gap of 100 bp  
\* 31215 38997: contig of 7783 bp in length  
\* 38998 39097: gap of 100 bp  
\* 39098 46783: contig of 7686 bp in length  
\* 46784 46883: gap of 100 bp  
\* 46884 54655: contig of 7772 bp in length  
\* 54656 54755: gap of 100 bp  
\* 54756 59749: contig of 4994 bp in length  
\* 59750 59849: gap of 100 bp  
\* 59850 63160: contig of 5311 bp in length  
\* 63161 65260: gap of 100 bp  
\* 65261 71291: contig of 6031 bp in length  
\* 71292 71391: gap of 100 bp  
\* 71392 75257: contig of 3866 bp in length  
\* 75258 75357: gap of 100 bp  
\* 75358 79204: contig of 3847 bp in length  
\* 79205 79304: gap of 100 bp  
\* 79305 83972: contig of 4668 bp in length  
\* 83973 84072: gap of 100 bp  
\* 84073 88891: contig of 4819 bp in length  
\* 88892 88991: gap of 100 bp  
\* 88992 93736: contig of 4745 bp in length  
\* 93737 93836: gap of 100 bp  
\* 93837 98281: contig of 4445 bp in length  
\* 98282 98381: gap of 100 bp  
\* 98382 102273: contig of 3892 bp in length  
\* 102274 102373: gap of 100 bp  
\* 102374 106147: contig of 3774 bp in length  
\* 106148 106247: gap of 100 bp  
\* 106248 110389: contig of 4142 bp in length  
\* 110390 110489: gap of 100 bp  
\* 110490 112888: contig of 2399 bp in length  
\* 112889 112988: gap of 100 bp  
\* 112989 115657: contig of 2669 bp in length  
\* 115658 115757: gap of 100 bp  
\* 115758 119437: contig of 3680 bp in length  
\* 119438 119537: gap of 100 bp  
\* 119538 123011: contig of 3474 bp in length  
\* 123012 123111: gap of 100 bp  
\* 123112 125411: contig of 2300 bp in length  
\* 125412 125511: gap of 100 bp  
\* 125512 128514: contig of 3003 bp in length  
\* 128515 128614: gap of 100 bp  
\* 128615 131919: contig of 3305 bp in length  
\* 131920 132019: gap of 100 bp  
\* 132020 133329: contig of 1310 bp in length  
\* 133330 133429: gap of 100 bp  
\* 133430 136197: contig of 2768 bp in length  
\* 136198 136297: gap of 100 bp  
\* 136298 138656: contig of 2359 bp in length  
\* 138657 138756: gap of 100 bp  
\* 138757 140802: contig of 2046 bp in length  
\* 140803 140902: gap of 100 bp  
\* 140903 142951: contig of 2049 bp in length  
\* 142952 143051: gap of 100 bp  
\* 143052 145259: contig of 2208 bp in length  
\* 145260 145359: gap of 100 bp  
\* 145360 147234: contig of 1875 bp in length  
\* 147235 147334: gap of 100 bp  
\* 147335 149339: contig of 2005 bp in length  
\* 149340 149439: gap of 100 bp  
\* 149440 150983: contig of 1544 bp in length  
\* 150984 151083: gap of 100 bp  
\* 151084 152215: contig of 1132 bp in length  
\* 152216 152315: gap of 100 bp  
\* 152316 153476: contig of 1161 bp in length  
\* 153477 153576: gap of 100 bp  
\* 153577 155210: contig of 1634 bp in length  
\* 155211 155310: gap of 100 bp  
\* 155311 156633: contig of 1323 bp in length  
\* 156634 156733: gap of 100 bp

\* 156734 157980: contig of 1247 bp in length  
\* 157981 158080: gap of 100 bp  
\* 158081 159424: contig of 1344 bp in length  
\* 159425 159524: gap of 100 bp  
\* 159525 160646: contig of 1122 bp in length  
\* 160647 160746: gap of 100 bp  
\* 160747 162107: contig of 1361 bp in length.

## FEATURES

source  
1..162107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18q11.2"  
/clone="RP11-863L8"  
1..12846  
/note="assembly\_fragment"  
12947..23222  
/note="assembly\_fragment"  
23323..31114  
/note="assembly\_fragment clone\_end:T7 vector\_side:right"  
31215..38997  
/note="assembly\_fragment"

Query Match . 68.4%; Score 21.2; DB 83; Length 162107;

Best Local Similarity 88.5%; Pred. No. 75;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 caggagacagaataaactgtggaata 30

Db 72809 CAGAGACAGAATAAATACTGCAGATA 72784

## RESULT 4

AC079751 163192 bp DNA HTG 10-SEP-2000

LOCUS Homo sapiens chromosome 19 clone RP11-61M18, WORKING DRAFT  
DEFINITION SEQUENCE, 19 unordered pieces.

AC079751

AC079751.1 GI:10047939

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Waterston, R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 163192)

Waterston, R.H.

Direct Submission

Submitted (10-SEP-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

## COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H.NH0061M18

----- Summary Statistics -----

Sequencing vector: M13; 100%

Chemistry: Dye-terminator Big Dye; 0% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 150903 bases at least Q40

Consensus quality: 155361 bases at least Q30

Consensus quality: 157169 bases at least Q20

Insert size: 165000; agarose-fp

Insert size: 161392; sum-of-contigs

Quality coverage: 4.38 in Q20 bases; agarose-fp

Quality coverage: 4.56 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs, their true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

|        |         |                              |
|--------|---------|------------------------------|
| 1      | 3640:   | contig of 3640 bp in length  |
| 3641   | 3740:   | gap of unknown length        |
| 3741   | 8087:   | contig of 4347 bp in length  |
| 8088   | 8187:   | gap of unknown length        |
| 8188   | 13561:  | contig of 5374 bp in length  |
| 13562  | 13661:  | gap of unknown length        |
| 13662  | 18899:  | contig of 5238 bp in length  |
| 18900  | 18999:  | gap of unknown length        |
| 19000  | 23802:  | contig of 4803 bp in length  |
| 23803  | 23902:  | gap of unknown length        |
| 23903  | 29983:  | contig of 6081 bp in length  |
| 29984  | 30083:  | gap of unknown length        |
| 30084  | 36841:  | contig of 6738 bp in length  |
| 36842  | 36941:  | gap of unknown length        |
| 36942  | 45404:  | contig of 8463 bp in length  |
| 45405  | 45504:  | gap of unknown length        |
| 45505  | 53225:  | contig of 8121 bp in length  |
| 53226  | 53725:  | gap of unknown length        |
| 53726  | 62720:  | contig of 8995 bp in length  |
| 62721  | 62820:  | gap of unknown length        |
| 62821  | 72763:  | contig of 9943 bp in length  |
| 72764  | 72863:  | gap of unknown length        |
| 72864  | 92764:  | contig of 19501 bp in length |
| 92765  | 92864:  | gap of unknown length        |
| 92865  | 114879: | contig of 21615 bp in length |
| 114880 | 114579: | gap of unknown length        |
| 114580 | 149140: | contig of 34561 bp in length |
| 149141 | 149240: | gap of unknown length        |
| 149241 | 150751: | contig of 1511 bp in length  |
| 150752 | 150851: | gap of unknown length        |
| 150852 | 153922: | contig of 3071 bp in length  |
| 153923 | 154022: | gap of unknown length        |
| 154023 | 156752: | contig of 2730 bp in length  |
| 156753 | 156852: | gap of unknown length        |
| 156853 | 159758: | contig of 2906 bp in length  |
| 159759 | 159758: | gap of unknown length        |
| 159859 | 163192: | contig of 3334 bp in length. |

FEATURES  
SOUND

```

1. 100132
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-61M18"
1. .3640
/note="assembly_name:Contig10"
3741. .8087
/note="assembly_name:Contig11"
8188. .13561
/note="assembly_name:Contig12"
13662. .18899
/note="assembly_name:Contig13"
19000. .23802
/note="assembly_name:Contig14"
23903. .29983
/note="assembly_name:Contig15"
30084. .36841
/note="assembly_name:Contig16"
36942. .45404
/note="assembly_name:Contig17"
45505. .53625
/note="assembly_name:Contig18"
53726. .62720
/note="assembly_name:Contig19"

```

|              |                                 |                                |
|--------------|---------------------------------|--------------------------------|
| misc_feature | 62821..72763                    | /note="assembly_name:Contig20" |
| misc_feature | 72864..92764                    | /note="assembly_name:Contig21" |
| misc_feature | 92865..114479                   | /note="assembly_name:Contig22" |
| misc_feature | 114580..149140                  | /note="assembly_name:Contig23" |
| misc_feature | 149241..150751                  | /note="assembly_name:Contig5"  |
| misc_feature | 150852..153922                  | /note="assembly_name:Contig6"  |
| misc_feature | 154023..156752                  | /note="assembly_name:Contig7"  |
| misc_feature | 156853..159758                  | /note="assembly_name:Contig8   |
|              | clone_end:SP6                   | vector_side:left"              |
| misc_feature | 159859..163192                  | /note="assembly_name:Contig9"  |
| BASE COUNT   | 45259 a 35953 c 35368 g 44806 t | 1806 others                    |
| ORIGIN       |                                 |                                |

```
Query Match 68.4%; Score 21.2; DB 76; Length 163192;
Best Local Similarity 88.5%; Pred. No. 75;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Qy 1 tcagcaggcacagaataaaaactgtgg 26  
||| ||| ||| ||| ||| ||| ||| |||  
77935 TCAGGAGGTACAGAAATAAACAGTGG 77960

RESULT 5  
AC079751/C

|            |   |           |     |     |             |
|------------|---|-----------|-----|-----|-------------|
| LOCUS      | AC079751  | 163192 bp | DNA | HTG | 10-SEP-2000 |
| DEFINITION | Homo sapiens chromosome 19 clone RP11-61M18, WORKING DRAFT SEQUENCE, 19 unordered pieces. |           |     |     |             |

ACCESSION AC079751  
VERSION AC079751.1 GI:10047939  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_L  
SOURCE human.

ORGANISM

REFERENCE  
1 (bases 1 to 163192)  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**AUTHORS**  
Waterston, R. H.  
(1988-1989)

**TITLE** The sequence of

JOURNAL  
Unpublished  
the sequence of homo sapiens clone

2 (bases 1 to 163193)

**AUTHORS**  
Waterston, R. H.  
(Duses, I. CO 103192)

| TITLE             | WATERSLOTT, R. H. |
|-------------------|-------------------|
| Direct submission |                   |

JOURNAL OF DIRECT SUBMISSIONS

Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0061M18
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Consensus program: Phrap; version 0.950319
Consensus quality: 150903 bases at least Q40
Consensus quality: 153361 bases at least Q30
Consensus quality: 157169 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 161392; sum-of-ctags

```

1

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L4257

Center clone name: 652-I-11

----- Summary Statistics

Sequencing vector: M13; M77815; 1% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 163169 bases at least Q40

Consensus quality: 170756 bases at least Q30

Insert size: 182000; agarose-fp

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

Quality coverage: 4.3 in Q20 ba.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 22 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 558: contig of 558 bp in length  
 \* 559 658: gap of 100 bp  
 \* 659 1793: contig of 1135 bp in length  
 \* 1794 1893: gap of 100 bp  
 \* 1894 3080: contig of 1187 bp in length  
 \* 3081 3180: gap of 100 bp  
 \* 3181 5062: contig of 1882 bp in length  
 \* 5063 5162: gap of 100 bp  
 \* 5163 7267: contig of 2105 bp in length  
 \* 7268 7367: gap of 100 bp  
 \* 7368 10092: contig of 2725 bp in length  
 \* 10093 10192: gap of 100 bp  
 \* 10193 14087: contig of 3895 bp in length  
 \* 14088 14187: gap of 100 bp  
 \* 14188 18512: contig of 4325 bp in length  
 \* 18513 18612: gap of 100 bp  
 \* 18613 23300: contig of 4688 bp in length  
 \* 23301 23400: gap of 100 bp  
 \* 23401 28085: contig of 4685 bp in length  
 \* 28086 28185: gap of 100 bp  
 \* 28186 32174: contig of 3989 bp in length  
 \* 32175 32274: gap of 100 bp  
 \* 32275 61458: contig of 29184 bp in length  
 \* 61459 61558: gap of 100 bp  
 \* 61559 68206: contig of 6648 bp in length  
 \* 68207 68306: gap of 100 bp  
 \* 68307 74318: contig of 6012 bp in length  
 \* 74319 74418: gap of 100 bp  
 \* 74419 82866: contig of 8448 bp in length  
 \* 82867 82966: gap of 100 bp  
 \* 82967 94336: contig of 11370 bp in length  
 \* 94337 94436: gap of 100 bp  
 \* 94437 107901: contig of 13465 bp in length  
 \* 107902 108001: gap of 100 bp  
 \* 108002 121825: contig of 13824 bp in length  
 \* 121826 121925: gap of 100 bp  
 \* 121926 138124: contig of 16199 bp in length  
 \* 138125 138224: gap of 100 bp  
 \* 138225 154039: contig of 15815 bp in length  
 \* 154040 154139: gap of 100 bp  
 \* 154140 172971: contig of 18832 bp in length  
 \* 172972 173071: gap of 100 bp  
 \* 173072 177748: contig of 4677 bp in length.  
 \* Location/Qualifiers  
 \* 1..177748  
 \* /organism="Homo sapiens"

FEATURES  
 source

/db\_xref="taxon:9606"  
 /chromosome="18"  
 /map="18"  
 /clone="RP11-652I11"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1..558

misc\_feature

/note="assembly\_fragment"

vector\_side:left"

clone\_end:SP6

659..1793

/note="assembly\_fragment"

1894..3080

/note="assembly\_fragment"

3181..5062

/note="assembly\_fragment"

5163..7267

/note="assembly\_fragment"

7368..10092

/note="assembly\_fragment"

10193..14087

/note="assembly\_fragment"

14188..18512

/note="assembly\_fragment"

18613..23300

/note="assembly\_fragment"

23401..28085

/note="assembly\_fragment"

28186..32174

/note="assembly\_fragment"

32275..61458

/note="assembly\_fragment"

61559..68206

/note="assembly\_fragment"

68307..74318

/note="assembly\_fragment"

74419..82866

/note="assembly\_fragment"

82967..94336

/note="assembly\_fragment"

94437..107901

/note="assembly\_fragment"

108002..121825

/note="assembly\_fragment"

121926..138124

/note="assembly\_fragment"

138225..154039

/note="assembly\_fragment"

154140..172971

/note="assembly\_fragment"

173072..177748

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

BASE COUNT 53384 a 35337 c 35096 g 51826 t 2105 others  
 ORIGIN

Query Match

Best Local Similarity 68.4%; Score 21.2; DB 64; Length 177748;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 cagggacagacagataaaactgtggaata 30

|||||

Db 143615 CAGAGACAGATAAACTGCAGATA 143640

RESULT 7

AC087677/c

LOCUS

DEFINITION

AC087677

ACCESSION

VERSION

AC087677 181041 bp DNA

Homo sapiens chromosome 18 clone RP11-863L8 map 18, WORKING DRAFT

SEQUENCE, 26 unordered pieces.

AC087677.2 GI:13123907

HTG

26-FEB-2001

Thu Oct 25 13:08:32 2001

us-09-462-955-1-copy\_941\_971.rge

```

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE 1 (bases 1 to 181041)
JOURNAL Birren, B., Linton, L., Nusbaum, C., and Lander, E.
REFERENCE Homo sapiens chromosome 18, clone RP11-863L8
AUTHORS Unpublished
2 (bases 1 to 181041)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barra, N., Bastien, V., Boguslavsky, L., Boukhaltier, B., Brown, A.,
Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Lakocque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienna, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 26, 2001 this sequence version replaced gi:12229452.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12275
Center clone name: 863.L.8
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172058 bases at least Q40
Consensus quality: 175838 bases at least Q30
Consensus quality: 177418 bases at least Q20
Insert size: 170000; agarose-1p
Insert size: 178541; sum-of-contigs
Quality coverage: 5.8 in Q20 bases; agarose-1p
Quality coverage: 5.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7738: contig of 7738 bp in length
* 7739 7838: gap of 100 bp
* 7839 8498: contig of 660 bp in length
* 8499 8598: gap of 100 bp
* 8599 9231: contig of 633 bp in length
* 9232 9331: gap of 100 bp
* 9332 9879: contig of 548 bp in length
*
* 9880 9979: gap of 100 bp
* 9980 10601: contig of 622 bp in length
* 10602 10701: gap of 100 bp
* 10702 12024: contig of 1323 bp in length
* 12025 12124: gap of 100 bp
* 12125 12920: contig of 796 bp in length
* 12921 13020: gap of 100 bp
* 13021 13844: contig of 824 bp in length
* 13845 13944: gap of 100 bp
* 13945 15024: contig of 1080 bp in length
* 15025 15124: gap of 100 bp
* 15125 38235: contig of 2311 bp in length
* 38236 38335: gap of 100 bp
* 38336 40005: contig of 1670 bp in length
* 40006 40105: gap of 100 bp
* 40106 42557: contig of 2452 bp in length
* 42558 42657: gap of 100 bp
* 42658 46488: contig of 3831 bp in length
* 46489 46588: gap of 100 bp
* 46589 52126: contig of 5538 bp in length
* 52127 52226: gap of 100 bp
* 52227 56744: contig of 4518 bp in length
* 56745 56844: gap of 100 bp
* 56845 62088: contig of 5244 bp in length
* 62089 62188: gap of 100 bp
* 62189 67648: contig of 5460 bp in length
* 67649 67748: gap of 100 bp
* 67749 74574: contig of 6826 bp in length
* 74575 74674: gap of 100 bp
* 74675 83911: contig of 9237 bp in length
* 83912 84011: gap of 100 bp
* 84012 96752: contig of 12741 bp in length
* 96753 96852: gap of 100 bp
* 96853 107204: contig of 10352 bp in length
* 107205 107304: gap of 100 bp
* 107305 123659: contig of 16355 bp in length
* 123660 123759: gap of 100 bp
* 123760 139462: contig of 15703 bp in length
* 139463 139562: gap of 100 bp
* 139563 160301: contig of 20739 bp in length
* 160302 160401: gap of 100 bp
* 160402 180220: contig of 19819 bp in length
* 180221 180320: gap of 100 bp
* 180321 181041: contig of 721 bp in length.
*
* Location/Qualifiers
* 1..181041
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="18"
* /map="18"
* /clone_lib="RP11-863L8"
* /clone="RP11-863L8"
* /note="assembly_fragment"
* 1..7738
* /note="assembly_fragment"
* /vector_side="left"
* 7839..8498
* /note="assembly_fragment"
* 8599..9231
* /note="assembly_fragment"
* 9332..9879
* /note="assembly_fragment"
* 9980..10601
* /note="assembly_fragment"
* 10702..12024
* /note="assembly_fragment"
* 12125..12920
* /note="assembly_fragment"
* 13021..13844
* /note="assembly_fragment"
* 13945..15024
* /note="assembly_fragment"
* 15125..38235

```



```

/misc_feature 38336 .40005 /note="assembly_fragment"
/misc_feature 40106 .42557 /note="assembly_fragment"
/misc_feature 42658 .46488 /note="assembly_fragment"
/misc_feature 46589 .52126 /note="assembly_fragment"
/misc_feature 52227 .56744 /note="assembly_fragment"
/misc_feature 56845 .62088 /note="assembly_fragment"
/misc_feature 62189 .67648 /note="assembly_fragment"
/misc_feature 67749 .74574 /note="assembly_fragment"
/misc_feature 74675 .83911 /note="assembly_fragment"
/misc_feature 84012 .96752 /note="assembly_fragment"
/misc_feature 96853 .107204 /note="assembly_fragment"
/misc_feature 107305 .123659 /note="assembly_fragment"
/misc_feature 123760 .139462 /note="assembly_fragment"
/misc_feature 139563 .160301 /note="assembly_fragment"
/misc_feature 160402 .180220 /note="assembly_fragment"
/misc_feature 180321 .181041 /note="assembly_fragment"
/note="assembly_fragment"
clone_end:T7
vector_side:right"

```

BASE COUNT 52864 a 37195 c 36606 g 51876 t 2500 others  
ORIGIN

Query Match 68.4%; Score 21.2; DB 77; Length 181041;  
Best Local Similarity 88.5%; Pred. No. 75;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 caggagacagaataaaactgtgaata 30  
||||| ||||| ||||| ||||| |||||

Db 94662 CAGAGACAGATAAACTGCAGAATA 94637

```

RESULT 8
AC023983 181239 bp DNA HTG 25-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-62116 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 35 unordered pieces.
ACCESSION AC023983
VERSION AC023983.3 GI:8076908
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181239)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-62116
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181239)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baidwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgaier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepli,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Liuers,T., Laroque,K., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Ollivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

#### TITLE JOURNAL

#### COMMENT

Direct Submission  
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7212065.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L6883  
Center clone name: 621\_L\_6  
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 160357 bases at least Q40  
Consensus quality: 170823 bases at least Q30  
Consensus quality: 174839 bases at least Q20  
Insert size: 183000; agarose-fp  
Quality coverage: 3.7 in Q20 bases; agarose-fp  
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 35 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1 1064: contig of 1064 bp in length
* 1065 1164: gap of 100 bp
* 1165 2658: contig of 1494 bp in length
* 2659 2758: gap of 100 bp
* 2759 4357: contig of 1599 bp in length
* 4358 4457: gap of 100 bp
* 4458 5644: contig of 1187 bp in length
* 5645 5744: gap of 100 bp
* 5745 7870: contig of 2126 bp in length
* 7871 7970: gap of 100 bp
* 7971 9676: contig of 1706 bp in length
* 9677 9776: gap of 100 bp
* 9777 11565: contig of 1789 bp in length
* 11566 11665: gap of 100 bp
* 11666 12026: contig of 361 bp in length
* 12027 12126: gap of 100 bp
* 12127 14427: contig of 2301 bp in length
* 14428 14527: gap of 100 bp
* 14528 17145: contig of 2618 bp in length
* 17146 17245: gap of 100 bp
* 17246 20711: contig of 3466 bp in length
* 20712 20811: gap of 100 bp
* 20812 23354: contig of 2543 bp in length
* 23355 23454: gap of 100 bp
* 23455 27289: contig of 3835 bp in length

```

[illegible]

feature lines that evidence is not experimental.

```

PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+ + + + +
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'AnnoMitter' (Hornischer & Bloecker).
+ + + + +
+ Programs used by 'AnnoMitter':
+ + + + +
> GeneFinder (Green), Vers. 084
> Organism: human
> Genscan (Burge & Karlin), Vers. 1.0
> Used matrix: vertebrate; Minimum score: 0
> Grail (Xu et al.), Vers. 1.3
> Organism: human
> Mzef (Zhang)
> Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
  & Skolnick)
> Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
  > BLASTN 2.0.9 (Altschul et al.)
  > Database(s): * RepBase (human), released 22-DEC-1995
  > * RepBase (primate), released 22-DEC-1995
  > Minimum score: 60;
  > 'ESTs': BLASTN 2.0.9 (Altschul et al.)
  > Database(s): * embl (EST), Vers. 60 (16-SEP-1999)
  > * emblnew (EST), Vers. 60+ (18-OCT-1999) Using unmasked
  sequence
  > Minimum score: 60; Minimum identity: 70 %;
  > 'GSSs': BLASTN 2.0.9 (Altschul et al.)
  > Database(s): * embl (GSS), Vers. 60 (16-SEP-1999)
  > * emblnew (GSS), Vers. 60+ (18-OCT-1999) Using unmasked
  sequence
  > Minimum score: 60; Minimum identity: 70 %;
  > 'Tandem Repeats': GDE 2.2 option 'tandem'
  > Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
  > Treat N's as mismatches? YES; Allow uniform consensi? NO >
  > 'Inverted Repeats': GDE 2.2 option 'inverted'
  > 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
  Islands': GDE 2.2 option 'cpG'
  > CpG island region size 100 bp;
  > Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
  e-PCR (Schuler)
  > Margin: 50; Number of mismatches allowed: 0; Word size: 7
  > STS database: 'dbSTS markers'
  > 'rRNA Scan': tRNAscan-SE (Low & Eddy), Vers. 1.11 Sequence
  overlaps with Acc.Nrs. AL050303 and AL078614.
FEATURES
source
1. 73465
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="RPICP704M13256Q2"
/map="21q11.1"
(3.461). (89.577)
/note="match: ESTs AL042464 R98070 T90747 AL044865 T90457
AA578081"
complement((240.666). (445.938))
/note="match: ESTs AF000593 R98316 AA584792 AT1761237
T83289 AJ003413"
341. .424
/note="homology = 72.6%, counts = 6"
/rpt_family="ggcctgatcagtg repeat"
/rpt_type=TANDEM
378. .449
/note="GENSCAN prediction, score = 6.32"
394. .420
/note="XPOUND prediction, score = 0.213"
649. .728
/note="GRAIL, score = 40%, comment = marginal"
complement((730.891). (923.1285))
/note="match: GSSs AQ341870 AQ074436 AQ079640 AQ437414
AQ475020"
826. .1098
/note="match: EST AL042124"
complement((6062.6213). .6315)
/note="match: ESTs AT372769 AA915976 AJ003450"

```

```

repeat_region
1523. 1795
/note="IRI, 85% complementary to IRI' (1523. 1795)"
/rpt_type=INVERTED
1523. 1795
/note="IRI', 85% complementary to IRI (826. 1098)"
/rpt_type=INVERTED
(1526.1986). (1734. 2126)
/note="match: GSSs AQ440604 AQ074436 B85689 AQ079640
AQ539584 AQ332219 AQ308154 AQ437414 AQ475020 AQ431870"
(1683.1819). (1795. 2126)
/note="match: ESTs AL044865 AA584792 AT1761237 AL040943
AL043508"
complement((1777.1781). (1870.1902))
/note="match: GSSs AQ610054 AQ428998"
2128. (2234.2312)
/note="match: ESTs AL043508 AL040943"
2128. (2238.2327)
/note="match: GSSs AQ308154 AQ539584 AQ332219"
2137. 2299
/note="GENSCAN prediction, score = 8.01"
2284. 2299
/note="XPOUND prediction, score = 0.356"
2328. 2560
/note="match: EST AA349997"
2328. (2408.2418)
/note="match: GSSs AQ332219 AQ539584"
complement(2462. .2553)
/note="MzEF prediction, score = 0.564"
(2587.2703). (3035.3037)
/note="match: ESTs H23104 N45556"
2687. 2870
/note="MzEF prediction, score = 0.709"
complement((2697.3121). (2939.3503))
/note="match: ESTs AA927218 AI970509 AJ003449 AL040461
AI216366 AA912892"
2929. 2944
/note="XPOUND prediction, score = 0.605"
3245. 3320
/note="homology = 76.3%, counts = 4"
/rpt_family="attggggatccaggca repeat"
(3438.3719). (3801.3933)
/note="match: ESTs AL042367 AL044856 AA316256 AL045430
AA385773 AL042369"
complement((3546.3822). (3845.3870))
/note="match: ESTs AW087494 AL040944 AI925550 AL040461
AA961620 AA878363 AI806762 N48708 AI015474 AI826553"
3653. 3870
/note="XPOUND prediction, score = 0.830"
3682. 3870
/note="GRAIL, score = 71%, comment = good"
4090. 4179
/note="match: EST 244176"
complement((4316.4583). (4561.4811))
/note="match: ESTs AW026376 AA864338 AL040944 AI125870
AI926636 AI126131 AI813897 AI925550 AL043509 AA970054
AI457652"
(4360.4364). (4415.4500)
/note="match: ESTs CI6613 AL044856 AL045430 AL042369
AL042367 AI372770"
complement(4367. .4500)
/note="match: GSS AQ188162"
complement(4512. 4661)
/note="GRAIL score = 71%, comment = good shadow"
complement(4542. 4693)
/note="match: GSS AQ188162"
(5664.5847). (5859.5940)
/note="match: ESTs AL046682 AL042369 AL042367 AA382548"
complement((5684.5922). (6164.6191))
/note="match: GSSs AQ0809119 AQ322681 AQ771684"
5997. 6248
/note="match: EST AL042124"
complement((6062.6213). .6315)
/note="match: ESTs AT372769 AA915976 AJ003450"

```

```

misc_feature      (6210.6239)..(6248.6428)
/notes="match: GSS AQ634548"
misc_feature      6759..6953
/notes="match: EST 244176"
exon              6925..7048
/notes="GENSCAN prediction, score = 0.54"
/notes="match: ESTs AL079705 R48206"
misc_feature      (7011.7306)..(7157.7478)
/notes="match: ESTs AL079705 R48206"
misc_feature      complement((7184.7229)..(7284.7298))
/notes="match: ESTs AI015474 AI145105 AI033019 AI222825
AI261513 AA915976 AJ003450 AI372769 AI674958 AA878363
AI806762 AI122718 N48708"
exon              7239..7298
/notes="XPOUND prediction, score = 0.503"
misc_feature      complement(7383..7452)
/notes="match: EST AL079706"
misc_feature      (7644.8094)..(7983.8186)
/notes="match: ESTs C00409 AA420700 AL040943 AL079705
AA420670"
misc_feature      complement((7681.8303)..(7930.8604))
/notes="match: ESTs AW087637 AA860908 AA961620 AA846418
AI216366 AI674958 AA878363 AI806762 Z40157 AI122718 N48708
AA93884 AA912892 AA927218 AW026376 AW087494 AI863316
AI145105 AA420670 AA778653 AI033019 AI222825 AI261513"

Query Match      67.7%; Score 21; DB 92; Length 73465;
Best Local Similarity 82.8%; Pred. No. 89;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggacagacataaaactgtgaata 30
||||| ||||| ||||| |||||
DB 43834 CAGCAGGTAAAGATAGATTAGGAATA 43862

RESULT 10
AL162851/C      152318 bp      DNA      HTG      23-JAN-2001
LOCUS      Homo sapiens chromosome 13 clone RP11-186M12, *** SEQUENCING IN
DEFINITION      PROGRESS ***, 28 unordered pieces.
ACCESSION      AL162851
VERSION      AL162851.5 GI:9930860
KEYWORDS      HTG: HTGS_PHASE1
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton,J.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9926516.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project information
Center project name: bA186M12
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 132981 bases at least Q40
Consensus quality: 140365 bases at least Q30
Consensus quality: 145091 bases at least Q20
Insert size: 149618; sum-of-contigs
Quality coverage: 177713; 5.3% error; agarose-fp
Quality coverage: 2.24x in Q20 bases; sum-of-contigs Quality
coverage: 2.25x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 2851 2950: contig of 2850 bp in length
* 2951 6292: contig of 3342 bp in length
* 6293 6392: gap of 100 bp
* 6393 9033: contig of 2641 bp in length
* 9034 9133: gap of 100 bp
* 9134 15105: contig of 5972 bp in length
* 15106 15205: gap of 100 bp
* 15206 21020: contig of 5815 bp in length
* 21021 21120: gap of 100 bp
* 21121 25318: contig of 4198 bp in length
* 25319 25418: gap of 100 bp
* 25419 28317: contig of 2899 bp in length
* 28318 28417: gap of 100 bp
* 28418 37932: contig of 9515 bp in length
* 37933 38032: gap of 100 bp
* 38033 40383: contig of 2351 bp in length
* 40384 40483: gap of 100 bp
* 40484 43573: contig of 3090 bp in length
* 43574 43673: gap of 100 bp
* 43674 52721: contig of 9048 bp in length
* 52722 52821: gap of 100 bp
* 52822 55850: contig of 3029 bp in length
* 55851 55950: gap of 100 bp
* 55951 64716: contig of 8766 bp in length
* 64717 64816: gap of 100 bp
* 64817 84621: contig of 19805 bp in length
* 84622 84721: gap of 100 bp
* 84722 87712: contig of 2991 bp in length
* 87713 87812: gap of 100 bp
* 87813 93264: contig of 5452 bp in length
* 93265 93364: gap of 100 bp
* 93365 100284: contig of 6920 bp in length
* 100285 100384: gap of 100 bp
* 100385 104611: contig of 4227 bp in length
* 104612 104711: gap of 100 bp
* 104712 107367: contig of 2656 bp in length
* 107368 107467: gap of 100 bp
* 107468 114106: contig of 6639 bp in length
* 114107 114206: gap of 100 bp
* 114207 118005: contig of 3799 bp in length
* 118006 118105: gap of 100 bp
* 118106 128798: contig of 10693 bp in length
* 128799 128898: gap of 100 bp
* 128899 131245: contig of 2347 bp in length
* 131246 131345: gap of 100 bp
* 131346 133727: contig of 2382 bp in length
* 133728 133827: gap of 100 bp
* 133828 135973: contig of 2046 bp in length
* 135974 144877: gap of 100 bp
* 144878 144977: contig of 8904 bp in length
* 144978 148859: contig of 3882 bp in length
* 148860 148959: gap of 100 bp
* 148960 152318: contig of 3359 bp in length.
*
FEATURES
      source
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="13"
      /clone="RP11-186M12"
      /clone_lib="RPCI-11.1"
      1..2850
      /note="assembly fragment:00886
      fragment_chain:1"
      2951..6292
      misc_feature
      misc_feature

```

|              |  |
|--------------|--|
| misc_feature | /note="assembly_fragment:00453<br>fragment_chain:1"<br>6393. .9033<br>/note="assembly_fragment:00046<br>fragment_chain:1"<br>9134. .15105<br>/note="assembly_fragment:00325<br>fragment_chain:1"<br>15206. .21020<br>/note="assembly_fragment:00973<br>fragment_chain:1"<br>21121. .25318<br>/note="assembly_fragment:00197<br>fragment_chain:2"<br>25419. .28317<br>/note="assembly_fragment:00937<br>fragment_chain:2"<br>28418. .37932<br>/note="assembly_fragment:00369<br>fragment_chain:2"<br>38033. .40383<br>/note="assembly_fragment:00014<br>fragment_chain:3"<br>40484. .43573<br>/note="assembly_fragment:00726<br>fragment_chain:3"<br>43674. .52721<br>/note="assembly_fragment:00081<br>fragment_chain:4"<br>52822. .55850<br>/note="assembly_fragment:00331<br>fragment_chain:4"<br>55951. .64716<br>/note="assembly_fragment:00876<br>fragment_chain:5"<br>64817. .84621<br>/note="assembly_fragment:00679<br>fragment_chain:5"<br>84722. .87712<br>/note="assembly_fragment:00003<br>87813. .93264<br>/note="assembly_fragment:00048<br>93365. .100284<br>/note="assembly_fragment:00088"<br>100385. .104611<br>/note="assembly_fragment:00101"<br>104712. .107367<br>/note="assembly_fragment:00142"<br>107468. .114106<br>/note="assembly_fragment:00256"<br>114207. .118005<br>/note="assembly_fragment:00310"<br>118106. .128798<br>/note="assembly_fragment:00335"<br>128999. .131245<br>/note="assembly_fragment:00511"<br>131346. .133727<br>/note="assembly_fragment:00657"<br>133828. .135873<br>/note="assembly_fragment:00664"<br>135974. .144877<br>/note="assembly_fragment:00779"<br>144978. .148859<br>/note="assembly_fragment:00823"<br>148960. .152318<br>/note="assembly_fragment:00373<br>clone_end:T7 |
|--------------|--|

|            | vector_side:right" |         |             |
|------------|--------------------|---------|-------------|
| BASE COUNT | 42602 a            | 31886 c | 31286 g     |
| ORIGIN     |                    | 43830 t | 2714 others |

### Query Match

67.7%; Score 21; DB 79; Length 152318;

```

Best Local Similarity      82.8%;    Pred. No. 91;
Matches   24; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Oy       2     cagcaggacagaataaaactgtggaata 30
           ||||| | | | | | | | | | | | | | | | | 
Db 28988 CAGCAGGTAAGAATACAAATTATGGAATA 28960

RESULT 11
HS214G14/c
LOCUS          HS214G14      168075 bp      DNA              08-FEB-2000
DEFINITION    Homo sapiens chromosome 21 clone RCIPT704G142L402 map 21qll.1, ***
SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION     AL050303
VERSION       AL050303.2 GI:6981890
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE     1 (bases 1 to 168075)
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS       Nordsiek,G., Conrad,A.), Dose,S., Grimm,M., Groet,J., Hornischer,K.,
               Loehnert,T.H., Nizetic,D., Scharfe.M., Schoen,O., Yaspo,M.L. and
               Bloecker,H.
TITLE         Direct Submission
JOURNAL       Submitted (25-MAY-1999) GBF, Dept. of Genome Analysis, Mascheroder
               Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
               nordsiek@gbf.de, bloecker@gbf.de
COMMENT       On Feb 16, 2000 this sequence version replaced gi:6966957.
               All annotations in this database entry are developed by
               computational tools. It is therefore not explicitly noted in the
               feature lines that evidence is not experimental.
               PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+++++++
+ * Analysis and annotation were performed with the automatic +
+ * 'first-pass' annotation and submission tool +
+ * 'AnnoMitter' (Hornischer & Bloecker). +
+ * Programs used by 'AnnoMitter': +
+++++++
> GeneFinder (Green), Vers. 084
+ * Organism: human
> GenScan (Burge & Karlin), Vers. 1.0
+ * Used matrix: vertebrate; Minimum score: 0
> Grail (Xu et al.), Vers. 1.3
+ * Organism: human
> Mzef (Zhang)
+ * Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
& Skolnick)
+ * Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
BLASTN 2.0.9 (Altschul et al.)
+ * Database(s): * RepBase (human), released 22-DEC-1995 .
+ * Rebase (primate), released 22-DEC-1995 .
RepBase (mammal), released 22-DEC-1995 .
+ * Minimum identity: 70 %; Minimum score: 60;
> 'EGTs': BLASTN 2.0.9 (Altschul et al.)
+ * Database(s): * embl (EST), Vers. 60 (16-SEP-1999) .
+ * emblnew (EST), Vers. 60+ (15-DEC-1999) . Using sequence with
masked repeats
+ * Minimum score: 60; Minimum identity: 90 %;
> 'GSSs': BLASTN 2.0.9 (Altschul et al.)
+ * Database(s): * embl (GSS), Vers. 60 (16-SEP-1999) .
+ * emblnew (GSS), Vers. 60+ (15-DEC-1999) . Using sequence with
masked repeats
+ * Minimum score: 60; Minimum identity: 90 %;
> 'tandem Repeats': GDE 2.2 option 'tandem'
+ * Minimum length 2 bp; Maximum length 20 bp; Score threshold 20 .
Treat N's as mismatches? YES; Allow uniform consensi? NO >
'inverted Repeats': GDE 2.2 option 'inverted'
> 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
Islands': GDE 2.2 option 'cpG'
+ * CpG island region size 100 bp;
+ * Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan';
e-PCR (Schuler)

```

```

/rpt_family="I1"
2201. .2293
/note="XPOND prediction, score = 0.707"
2331. .2820
/note="89% identity: matches 2701. .3192 of consensus"
2865. .2995
/rpt_family="L1"
/note="IR5, 85% complementary to IR5' (.4561. .4691)"
/rpt_type=INVERTED
2872. .2995
/note="IR4, 87% complementary to IR4' (.4561. .4684)"
/rpt_type=INVERTED
2875. .2995
/note="IR2, 82% complementary to IR2' (.3283. .3403)
IR3, 89% complementary to IR3' (.4561. .4681)"
/rpt_type=INVERTED
complement((2330.2934). (2983.2990))
/note="match: GSSS AQ471711 AQ041712 AQ0410967 AQ034367
AQ219103 B55537 AQ185675 AQ19038"
(2933.2947). (.2379.3025)
/note="match: ESTS AW188087 AA577992 A1979227 AF623755
AA423972 AL048971 T35301 AI659905 AI927405 F01778 AI963294
AA883246 AA824522 AA932573 AI1199402 AT708022 AI300484
AA931704 AI697170 AA448484 Z38954 AI079687 AI436442 FO1552
A2226611 AA641060 AI191577 AA962425 AA989247 AI341212
AA657489 AI799492 AT247389 AA911085 AI979195 AA770179
AI867437 AI798417 AA262248 AA987270 AI458178 AI701161
AW193871"
misc_feature
complement(2993. .2987)
/note="match: ESTS AA226235 AA432049"
2986. .3024
/note="match: GSS AQ041712"
complement((3067.314).(.3192)
/note="match: GSSS AQ207681 AQ185675 B30835"
3152. .3192
/note="match: GSS B18150"
(3246.3308). (.3300.3354)
/note="match: GSSS AQ471711 AQ027470 AQ672190 AQ007282
B30835"
(3283.3291). (.3345.3354)
/note="match: ESTS AA226235 AA318324 AA432049"
complement((3283.3299). (3344.3351))
/note="match: ESTS AW188087 T02820 A1979227 AF623755
AA423972 AL048971 T35301 AI659905 AI927405 F01778 AI963294
AA883246 AA824522 AA932573 AI1199402 AT708022 AI300484
AA931704 AI697170 AA448484 Z38954 AI079687 AI436442 FO1552
A2226611 AA641060 AI191577 AA962425 AA989247 AI341212
AA657489 AI799492 AT247389 AA911085 AI979195 AA770179
AI867437 AI798417 AA262248 AA987270 AI458178 AI701161
AW193871"
repeat_region
3283. .3403
/note="IR2 , 82% complementary to IR2 (.2875. .2995)"
/rpt_type=INVERTED
3401. .3502
/note="match: EST T02820"
complement((3445.3569). (.3524.3623))
/note="match: GSSS AQ562833 AQ567848 AQ185675 B30835
AQ041712 AQ410967"
(3571.3572). (3612.3625)
/note="match: ESTS AW188087 A1979227 AF623755 AA423972
AL048971 T35301 AI659905 AI927405 F01778 AI963294 AA883246
AA824522 AA932573 AI1199402 AT708022 AI300484 AA931704
AI697170 AA448484 Z38954 AI079687 AI436442 FO1552 AA226611
AA641060 AI191977 AA962425 AA989247 AI341212 AA657489
FO1778"
Query Match 67.7%; Score 21; DB 84; Length 168075;
Best Local Similarity 82.8%; Pred.No. 92;
... .. Mismatches 0; ... Indels 5; Gaps 1

```

```

RESULT 12
HS21C004/c
LOCUS      340000 bp      DNA
DEFINITION Homo sapiens chromosome 21 segment HS21C004.
ACCESSION AL163204 AP001659 BA000005
VERSION    AL163204.2 GI:7717247
KEYWORDS
SOURCE
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 340000)
AUTHORS    Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
            Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
            Direct Submission
            Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
            The Chromosome 21 Mapping and Sequencing Consortium consists of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagami-hara 228-8555, Japan.
            * e-mail: sakaki@gsr.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e-mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Dept. of Molecular Biology, *
            Tokyo 160-8582, Japan.
            * e-mail: shimizu@mb.med.keio.ac.jp
            * URL: http://adenine.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e-mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/.
            Location/Qualifiers
            1..340000
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q11.1"
            <1..3300
            /note="Accession No. AP001465"
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q11.1"
            /clone="R663H15, 5' partial"
            /clone_lib="RPC1-11 BAC library"
            473..1061
            repeat_region

FEATURES
source
    /note="L1PA12"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    complement(1115..1235)
    /note="L1PA13"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    1228..2242
    /note="L1PA12"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    2243..2552
    /note="Alu"
    /rpt_family="SINE/Alu"
    /rpt_type=DISPERSED
    2301..170825
    /note="Accession No. AL050303"
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q11.1"
    /clone="P214G14"
    /clone_lib="RPC11,3-5 PAC library"
    2553..2567
    /note="L1PA12"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    2569..3516
    /note="MER4-internal"
    /rpt_family="LFR/MER4-group"
    /rpt_type=DISPERSED
    complement(3535..4595)
    /note="L1PA7"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    4622..5134
    /note="L1P"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    complement(9593..9894)
    /note="AluJo"
    /rpt_family="SINE/Alu"
    /rpt_type=DISPERSED
    10899..11367
    /note="L1M1"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    11506..11811
    /note="AluSg"
    /rpt_family="SINE/Alu"
    /rpt_type=DISPERSED
    complement(11883..12556)
    /note="L1MEC"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    complement(12902..13457)
    /note="L1MCC"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    13458..13487
    /note="(GC)n"
    /rpt_family="Simple_repeat"
    /rpt_type=TANDEM
    complement(13488..13708)
    /note="L1MCC"
    /rpt_family="LINE/L1"
    /rpt_type=DISPERSED
    complement(14935..15084)
    /note="MIR"

```

Thu Oct 25 13:08:32 2001

```

/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
15303..15409
/note="(TTC)n"
/rpt_type=TANDEM
complement(15413..15735)
/note="AluJb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(16256..16379)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(16626..16919)
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(17158..17447)
/note="AluSq"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(17603..17794)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
18054..18154
/note="MLT2H"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
complement(18708..19019)
/note="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
19375..19481
/note="AluSg1"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(19492..19968)
/note="LIMB3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
20554..21037
/note="LIM4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
21497..21749
/note="LTR37B"
/rpt_family="LTR/MBR4-group"
/rpt_type=DISPERSED
23106..23276
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
23296..23470
/note="(CA)n"
/rpt_family="Simple_repeat"

Query Match 67.78; Score 21; DB 92; Length 340000;
Best Local Similarity 82.8%; Pred.No. 94;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggaacagataaaactgtggaata 30
||||| ||||| ||||| ||||| |||||
Db 166036 CAGCAGGTAAGAAATAGAAATATGGAATA 166008

RESULT 13
AC021220 214348 bp DNA HPG 15-NOV-2000
LOCUS Homo sapiens chromosome 4 clone RP11-530I17, WORKING DRAFT
DEFINITION
SEQUENCE, 38 unordered pieces.
ACCESSION AC021220

```

VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AC021220.4 GI:11178158  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 214348)  
 Waterston,R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 214348)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (15-JAN-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Nov 15, 2000 this sequence version replaced gi:9309528.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H\_NH0530I17  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 87%  
 Sequencing vector: plasmid; 13%  
 Chemistry: Dye-primer ET; 87% of reads  
 Chemistry: Dye-terminator Big Dye; 13% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 195065 bases at least Q40  
 Consensus quality: 200600 bases at least Q30  
 Consensus quality: 203756 bases at least Q20  
 Insert size: 161000; agarose-fp  
 Insert size: 210648; sum-of-contigs  
 Quality coverage: 4.96 in Q20 bases; agarose-fp  
 Quality coverage: 4.62 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1584: contig of 1584 bp in length  
 \* 1585 1684: gap of unknown length  
 \* 1685 4194: contig of 2510 bp in length  
 \* 4195 4294: gap of unknown length  
 \* 4295 6424: contig of 2130 bp in length  
 \* 6425 6524: gap of unknown length  
 \* 6525 9844: contig of 3320 bp in length  
 \* 9845 9945: gap of unknown length  
 \* 9945 14101: contig of 4157 bp in length  
 \* 14102 14201: gap of unknown length  
 \* 14202 21279: contig of 7078 bp in length  
 \* 21280 21346: contig of 9967 bp in length  
 \* 21347 31347: gap of unknown length  
 \* 31347 40526: contig of 9080 bp in length  
 \* 40527 40626: gap of unknown length  
 \* 40627 52862: contig of 12236 bp in length  
 \* 52863 52963: contig of 13128 bp in length  
 \* 52963 66091: gap of unknown length  
 \* 66091 82659: contig of 16459 bp in length  
 \* 82660 82759: gap of unknown length  
 \* 82760 104797: contig of 22038 bp in length  
 \* 104798 104897: gap of unknown length  
 \* 104898 129319: contig of 24422 bp in length  
 \* 129320 129419: gap of unknown length  
 \* 129420 176511: contig of 47092 bp in length



```

* 176512 176611: gap of unknown length
* 176612 177666: contig of 1055 bp in length
* 177667 177666: gap of unknown length
* 177667 178997: contig of 1231 bp in length
* 178998 179097: gap of unknown length
* 179098 180242: contig of 1145 bp in length
* 180243 180342: gap of unknown length
* 180343 181442: contig of 1100 bp in length
* 181443 181542: gap of unknown length
* 181543 182642: contig of 1100 bp in length
* 182643 182742: gap of unknown length
* 182743 184165: contig of 1423 bp in length
* 184166 184266: gap of unknown length
* 184267 185612: contig of 1247 bp in length
* 185613 185612: gap of unknown length
* 185613 187121: contig of 1509 bp in length
* 187122 188524: contig of 1303 bp in length
* 188525 188625: gap of unknown length
* 188626 189782: contig of 1158 bp in length
* 189783 189882: gap of unknown length
* 189883 191312: contig of 1430 bp in length
* 191313 191412: gap of unknown length
* 191413 192837: contig of 1425 bp in length
* 192838 192937: gap of unknown length
* 192938 194117: contig of 1180 bp in length
* 194118 194217: gap of unknown length
* 194218 195948: contig of 1731 bp in length
* 195949 196048: gap of unknown length
* 196049 197744: contig of 1696 bp in length
* 197745 197844: gap of unknown length
* 197845 199489: contig of 1645 bp in length
* 199490 199589: gap of unknown length
* 199590 201304: contig of 1715 bp in length
* 201305 201404: gap of unknown length
* 201405 203316: contig of 1912 bp in length
* 203317 203416: gap of unknown length
* 203417 204490: contig of 1074 bp in length
* 204491 204590: gap of unknown length
* 204591 206677: gap of unknown length
* 206678 206776: contig of 2086 bp in length
* 206777 208803: contig of 2027 bp in length
* 208804 208903: gap of unknown length
* 208904 210597: contig of 1694 bp in length
* 210598 210698: gap of unknown length
* 210699 212097: contig of 1400 bp in length
* 212098 212197: gap of unknown length
* 212198 214348: contig of 2151 bp in length.
FEATURES
    source
    1. .214348
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="4"
    /clone="RP11-530117"
    1. .1584
    /note="assembly_name:Contig100"
    1685. .4194
    /note="assembly_name:Contig101"
    4295. .6424
    /note="assembly_name:Contig102"
    6525. .9844
    /note="assembly_name:Contig103"
    9945. .14101
    /note="assembly_name:Contig104"
    14202. .21279
    /note="assembly_name:Contig105"
    21380. .31346
    /note="assembly_name:Contig106"
    31447. .40526
    /note="assembly_name:Contig107"
    vector_side:left"
    40627. .52862

```

```

misc_feature /note="assembly_name:Contig108"
52963. .66090
misc_feature /note="assembly_name:Contig109"
66191. .82659
misc_feature /note="assembly_name:Contig110"
82760. .104797
misc_feature /note="assembly_name:Contig111"
104898. .129319
clone_end:SP6
vector_side:right"
129420. .176511
misc_feature /note="assembly_name:Contig113"
176612. .177666
misc_feature /note="assembly_name:Contig133"
177667. .178997
misc_feature /note="assembly_name:Contig37"
179098. .180242
misc_feature /note="assembly_name:Contig56"
180343. .181442
misc_feature /note="assembly_name:Contig59"
181543. .182642
misc_feature /note="assembly_name:Contig66"
182743. .184165
misc_feature /note="assembly_name:Contig69"
184266. .185512
misc_feature /note="assembly_name:Contig75"
185613. .187121
misc_feature /note="assembly_name:Contig77"
187222. .188524
misc_feature /note="assembly_name:Contig78"
188625. .189782
misc_feature /note="assembly_name:Contig80"
189883. .191312
misc_feature /note="assembly_name:Contig85"
191413. .192837
misc_feature /note="assembly_name:Contig86"
192938. .194117
misc_feature /note="assembly_name:Contig87"
194218. .195948
misc_feature /note="assembly_name:Contig88"
196049. .197744
misc_feature /note="assembly_name:Contig89"
197845. .199489
misc_feature /note="assembly_name:Contig90"
199590. .201304
misc_feature /note="assembly_name:Contig91"
201405. .203316
misc_feature /note="assembly_name:Contig93"
203417. .204490

```

Query Match 67.18; Score 20.8; DB 66; Length 214348;  
 Best Local Similarity 91.7%; Pred. No. 1.le+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cagcaggacagataaaactgtg 25  
 ||||| ||||| ||||| |||||  
 Db 55931 CACCAGGAGAAAGATAAACTGTG 55954

RESULT 14  
 AB02952511  
 LOCUS AB02952511 1211 bp DNA ROD 14-APR-2000  
 DEFINITION Mus musculus Mea2/Golga3 gene, exon 12 and 13.  
 ACCESSION AB029530  
 VERSION AB029530.1 GI:6429304  
 KEYWORDS male enhanced antigen 2/golgi autoantigen, golgin subfamily a, 3;  
 Mea2/Golga3.  
 11 of 18  
 SEGMENT Mus musculus (sub\_species:domesticus, strain:C57BL/6) DNA.  
 SOURCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

CC hypothetical: No;
CC anti-sense: No;
FH key Location/Qualifiers
FT source 1. .4621
FT /organism='Mus musculus domesticus' FT
FT /tissue_type='Testis'
FT l<. .4621
FT /gene='Mea-2' .
FT Location/Qualifiers
FT source 1. .4621
FT /organism='Mus musculus'
FT /db_xref="taxon:10090"
FT 1384 a 1058 c 1306 g 872 t 1 others
FEATURES
source
BASE COUNT 1384 a 1058 c 1306 g 872 t 1 others
ORIGIN

Query Match 66.5%; Score 20.6; DB 10; Length 4621;
Best Local Similarity 85.2%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cagcaggacagataaaaactgtggaa 28
| | | | | | | | | | | | | | | |
Db 2739 CTGCAGGGGAGAGAGAACTGTGGAA 2765

Search completed: October 24, 2001, 11:42:34
Job time: 6339 sec

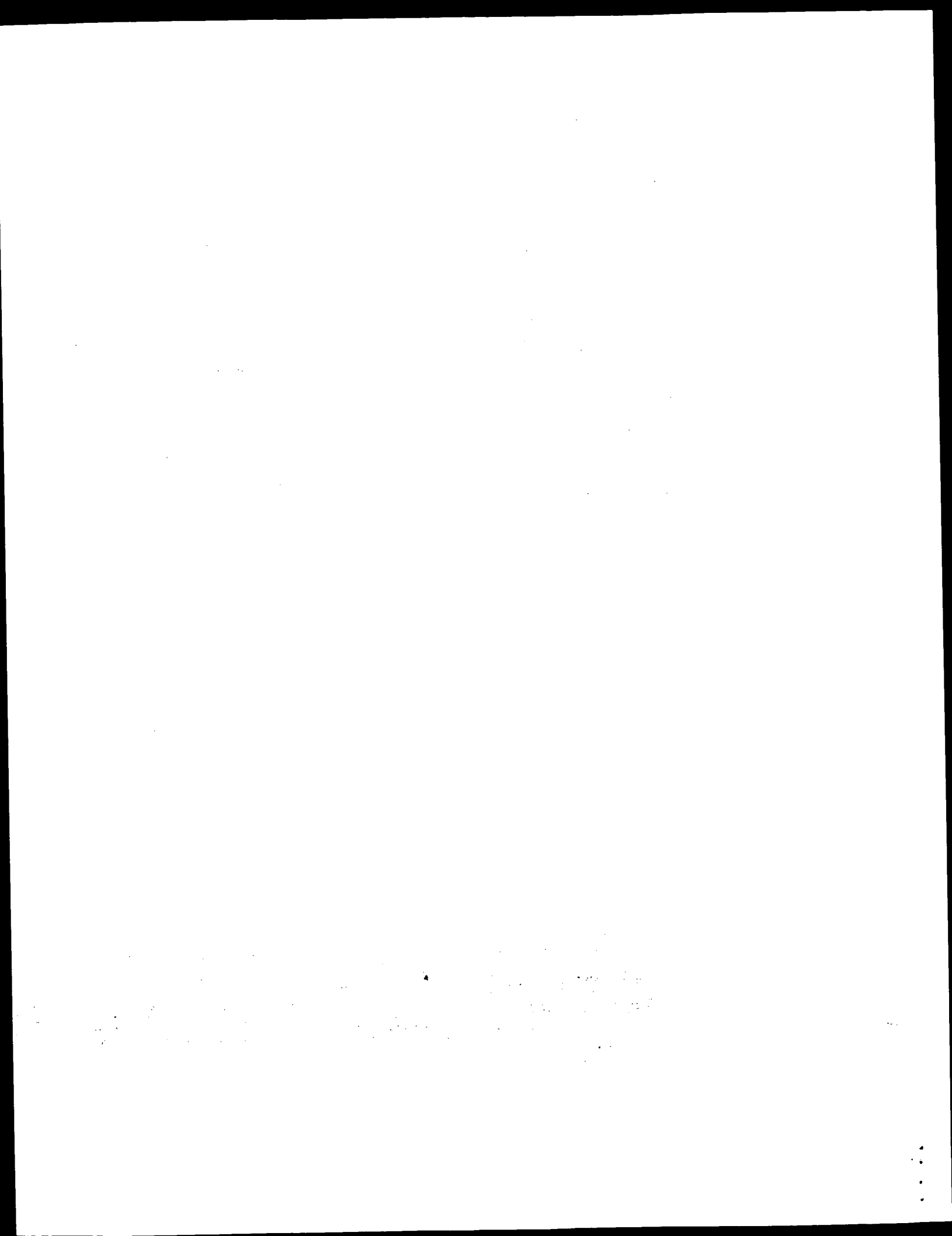
```

..  
Thu Oct 25 13:08:32 2001

us-09-462-955-1\_copy\_941\_971.rge

---

Page 19



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:06:56 ; Search time 393.52 Seconds  
(without alignments)  
49.464 Million cell updates/sec

Title: US-09-462-955-1-copy\_941\_971

Perfect score: 31

Sequence: 1 tcagcaggacagataaaactgtggaatat 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 20.6  | 66.5        | 4621   | AA74034  | Mouse male-enhance |
| 2          | 20.6  | 66.5        | 4621   | AA74034  | Mouse male enhance |
| 3          | 19.8  | 63.9        | 232    | AA60888  | Human IRG27 relate |
| 4          | 19.4  | 62.6        | 20844  | AA81460  | N. meningitidis pa |
| 5          | 19.4  | 62.6        | 349980 | AA72161  | Neisseria meningit |
| 6          | 19.4  | 62.6        | 837096 | AA81489  | N. meningitidis pa |
| 7          | 19.2  | 61.9        | 802    | AA44457  | Human secreted exp |
| 8          | 19    | 61.3        | 360    | AA65352  | Novel human polynu |
| 9          | 19    | 61.3        | 404    | AA65328  | Novel human polynu |
| 10         | 19    | 61.3        | 5578   | AA76023  | TGF-beta-2 promote |
| 11         | 18.8  | 60.6        | 476    | AA254436 | Neisseria meningit |

|   |    |      |      |        |    |          |                    |
|---|----|------|------|--------|----|----------|--------------------|
| c | 12 | 18.8 | 60.6 | 526    | 21 | AA254437 | Neisseria meningit |
| c | 13 | 18.8 | 60.6 | 1001   | 21 | AA254437 | Neisseria meningit |
|   | 14 | 18.8 | 60.6 | 1325   | 21 | AA254437 | Neisseria meningit |
|   | 15 | 18.8 | 60.6 | 2061   | 20 | AA254437 | Neisseria meningit |
|   | 16 | 18.8 | 60.6 | 2373   | 17 | AA254437 | Neisseria meningit |
|   | 17 | 18.8 | 60.6 | 2376   | 17 | AA254437 | Neisseria meningit |
|   | 18 | 18.8 | 60.6 | 2379   | 17 | AA254437 | Neisseria meningit |
| c | 19 | 18.8 | 60.6 | 2379   | 17 | AA254437 | Neisseria meningit |
| c | 20 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
|   | 21 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
|   | 22 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 23 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 24 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 25 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 26 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 27 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 28 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 29 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 30 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 31 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 32 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 33 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 34 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 35 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 36 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 37 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 38 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 39 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 40 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 41 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 42 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 43 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 44 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |
| c | 45 | 18.8 | 60.6 | 349980 | 21 | AA254437 | Neisseria meningit |

ALIGNMENTS

|         |  |
|---------|--|
| RESULT  | 1  |
| AA74034 |  |
| ID      | AA74034 standard; cDNA: 4621 BP.             |
| XX      |  |
| AC      |  |
| AC      |  |
| XX      |  |
| XX      |  |
| DF      | 16-SEP-1997 (first entry)                    |
| XX      |  |
| DE      | Mouse male-enhanced antigen-2 encoding cDNA. |
| XX      |  |
| XX      |  |
| KW      | Mouse; MEA-2; detecting mutation; ds.        |
| XX      |  |
| OS      | Mus musculus domesticus.                     |
| XX      |  |
| FH      | Key  |
| CDS     | Location/Qualifiers                          |
| FT      | 570..4547                                    |
| FT      | /*tag= a                                     |
| FT      | /product= Male-enhanced_antigen-2            |
| FT      | misc_difference 307                          |
| FT      | /*tag= b                                     |
| FT      | /note= "Unspecified"                         |
| XX      |  |
| PN      | JP09121869-A.                                |
| XX      |  |
| PD      | 13-MAY-1997.                                 |
| XX      |  |
| PF      | 07-NOV-1995; 95JP-0311638.                   |
| XX      |  |
| PR      | 07-NOV-1995; 95JP-0311638.                   |
| XX      |  |
| PA      | (ITOH-) ITO HAM KK.                          |
| XX      |  |
| DR      | WPI: 1997-314229/29.                         |
| DR      | P-PSDB; AAW19540.                            |

XX Male-enhanced antigen Mea-2 gene - especially from mouse, useful for  
 PT detecting mutation(s)  
 XX Claim 3; Page 7-9; 13pp; Japanese.  
 PS The present sequence encodes male-enhanced antigen-2 (Mea-2), which  
 CC has been derived from a mouse. The polynucleotide can be  
 CC used for the detection of mutations affecting the Mea-2 gene.  
 XX Sequence 4621 BP; 1384 A; 1058 C; 1306 G; 872 T; 1 other;  
 SQ

Query Match 66.5%; Score 20.6; DB 18; Length 4621;  
 Best Local Similarity 85.2%; Pred. No. 16;  
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 cagcaggacagaataaaactgtggaa 28  
 | ||||| |||| || |||||  
 Db 2739 ctgcaggagagagaagaactgtggaa 2765

RESULT 2  
 AAX04132  
 ID AAX04132 standard; cDNA; 4621 BP.  
 XX  
 AC AAX04132;  
 XX  
 DT 14-APR-1999 (first entry).  
 XX  
 DE Mouse male enhanced antigen 2 encoding cDNA.  
 XX  
 KW Mouse; male enhanced antigen 2; Mea-2; Mus musculus domesticus;  
 KW spermatogenesis; regulation; contraceptive; sterile; inhibition; ds.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 570..4547  
 FT /\*tag= a  
 XX  
 JP11018622-A.  
 XX  
 PD 26-JAN-1999.  
 XX  
 PF 04-JUL-1997; 97JP-0179490.  
 XX  
 PR 04-JUL-1997; 97JP-0179490.  
 XX  
 PA (ITOH-) ITO HAM KK.  
 XX  
 DR WPI; 1999-160962/14;  
 DR P-PSDB; AAW94391.  
 XX  
 XX Regulation of spermatogenesis using Mea-2 gene information - using  
 PT anti-sense oligo- or poly:nucleotide(s), used for production of  
 PT contraceptives  
 XX  
 PS Claim 2; Page 8-12; 27pp; Japanese.  
 XX  
 CC The present sequence encodes mouse male enhanced antigen 2 (Mea-2). The  
 CC present invention describes the regulation of spermatogenesis by using  
 CC Mea-2 information. A non-human living organism can have its  
 CC spermatogenesis inhibited by breakage of the whole or part of the Mea-2  
 CC gene. Also described are: (1) the creation of the spermatogenesis-  
 CC inhibited organism; (2) a drug composition containing an oligonucleotide  
 CC or polynucleotide containing base sequences that pair with at least part  
 CC of the Mea-2 gene and are able to inhibit the expression of Mea-2 gene;  
 CC and (3) the creation of an aimed gene-possessing organism using the  
 CC spermatogenesis inhibited organism. The organism is useful for producing  
 CC contraceptive drugs.  
 XX  
 SQ Sequence 4621 BP; 1384 A; 1059 C; 1306 G; 872 T; 0 other;

Query Match 66.5%; Score 20.6; DB 20; Length 4621;  
 Best Local Similarity 85.2%; Pred. No. 16;  
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 cagcaggacagaataaaactgtggaa 28  
 | ||||| |||| || |||||  
 Db 2739 ctgcaggagagagaagaactgtggaa 2765

RESULT 3  
 AAC60888  
 ID AAC60888 standard; DNA; 232 BP.  
 XX  
 AC AAC60888;  
 XX  
 DT 07-FEB-2001 (first entry)  
 XX  
 DE Human IRG27 related DNA sequence SEQ ID NO:29.  
 XX  
 KW Immortalised related gene 27; IRG27; p53; carcinostatic factor;  
 KW diagnosis; cancer; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000060073-A1.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 23-MAR-2000; 2000WO-JP01796.  
 XX  
 PR 31-MAR-1999; 99JP-0093641.  
 XX  
 PA (SUMU) SUMITOMO PHARM CO LTD.  
 XX  
 PI Enjoji T, Tohdoh N;  
 XX  
 DR WPI; 2000-638465/61.  
 XX  
 PT Diagnostic agent comprising sequences from the gene encoding the  
 PT peptide IRG27 expressed by p53 inactivation, useful for the diagnosis  
 PT of a broad range of cancers  
 XX  
 PS Claim 2; Page 76; 90pp; Japanese.  
 XX  
 CC The present invention describes a cancer diagnostic agent which contains  
 CC sequences of at least 17 bases derived from the sense or antisense chain  
 CC of the gene encoding the immortalisation related gene 27 (IRG27) protein,  
 CC a protein whose expression is induced by the inactivation of the  
 CC carcinostatic factor p53 and whose expression is elevated in cancer  
 CC tissue. Also described are: (1) antibodies to peptides encoded by the  
 CC DNA; (2) a method for cancer diagnosis using the agents; and (3) genes  
 CC whose expression is induced by the inactivation of p53. The method is  
 CC useful for the diagnosis of a broad range of cancers using the DNA  
 CC sequences or antibodies, including cancer of the stomach, oesophagus,  
 CC liver, womb, breast, bladder, ureter, kidney, thyroid, ovary and  
 CC fallopian tubes. The present sequence represents an IRG27 related DNA  
 CC sequence which is given in the present invention.  
 XX  
 SQ Sequence 232 BP; 86 A; 42 C; 41 G; 63 T; 0 other;

Query Match 63.9%; Score 19.8; DB 21; Length 232;  
 Best Local Similarity 77.4%; Pred. No. 20;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 tcagcaggacagaataaaactgtggaat 31  
 | ||||| |||| || |||||  
 Db 69 tcagcagataaagagataaaactgtggaacat 99

RESULT 4

[illegible]

```
Query Match      62.6%; Score 19.4; DB 21; Length 837096;
Best Local Similarity 79.3%; Pred. No. 1.1e+02;
Matches 23: Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

Qy 2 cagcaggacagaataaaactgtggaata 30  
||||| ||| | | | |  
306539 CAGCAGGGACATAAATACACGGTTGAAGA 306511

## 7. RESULT

AAA44457  
ID AAA4457 standard: cDNA: 802 bp.

XX  
AC AAA44457:

21-AUG-2000 (first entry)

human secreted expressed sequence tag SEO ID NO:1032.

Human; mouse; chicken; rat; secreted expressed sequence tag; SST; expressed sequence tag; EST; probe; chemotactic; proliferative; immunomodulatory; haematopoietic; chemokineti; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antiadrenergic; antiasthmatic; vulnerrary; antiparkinsonian; anticancer; osteopathic; neuroprotective; nootropic; antipsoriatic; cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine; autoimmune disorder; multiple sclerosis; allergic condition; insulin depend diabetes; asthma; myeloid cell deficiency; ulcer; lymphoid cell deficiency; burn; osteoporosis; osteoarthritis; central nervous system disorder; Alzheimer's disease; stroke; parkinson's disease; Huntington's disease; coagulation disorder; haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour; infection; depression; psoriasis; ss.

XX  
US  
OS  
Homo sapiens.

XX  
PN  
W0200021991-A1.XX  
20-APR-2000-

XX  
PF  
15-OCT-1999: 99WO-US24206.

XX  
15-OCT-1998  
98UTS-0104436-XX  
F3 /CEMV \ GENETICS INST INC  
XX[illegible]

PI Merberg D, Treacy M, Bowman MR;

DR WPI; 2000-317938/27.

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (sESTs), useful for treating various disorders such as autoimmune infectious, and central nervous system disorders -

PS Claim 1: Page 467; 803pp; English.

AAA43426 to AAA45925 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnery; antiulcer; osteopathic; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions

2 cagcaggacagaaataaaaactgtggaata 30

## RESULT 6

AAA81489/c  
ID AAA81489 standard: DNA: 837096 BP.

XX  
AC  
AAA81489.

XX  
DT 04-DEC-2000 (first entry)[illegible]

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B: MenB: ds.  
KW

XX  
Neisseria meningitidis.

XX  
DN  
W0200022430-A2

XX  
DD  
20-APP-2000

08-05T-1000- 99W0-IIS23573

XX  
DD  
00-000-1000. 0000-0103794

PR 30-APR-1999; 9905-0132068.

PA (CHIR ) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettellin H, Veneri JC;  
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizzia M;

XX  
DP  
WPT: 2000-318079/27.

Isolated nucleotide sequences of *Neisseria meningitidis* which can be used in the diagnosis and treatment of *N. meningitidis* infection and other *Neisseria* infections for example, *N. gonorrhoea*.

XX 620-865. 1760nn. English

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

XX  
sequence 837096 BP: 207534 A: 227065 C: 205215 G: 197280 T: 2 other;



CC (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal or viral infections, depression and  
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
 CC in the exemplification of the present invention.

XX Sequence 802 BP; 270 A; 145 C; 170 G; 216 T; 1 other;

Query Match 61.9%; Score 19.2; DB 21; Length 802;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 8 ggacagaataaaactgtggaatat 31  
 || ||||| ||||| ||||| |||||  
 Db 668 gggcagaataaaacagtgcgaatat 691

RESULT 8  
 AAF65352  
 ID AAF65352 standard; cDNA; 360 BP.

XX AAF65352;

DT 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 1108.

DE Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
 KW breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a  
 PT mammalian cell and detecting cancer, particularly of the colon or  
 PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 703; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to produce probes for detection of mRNA and to produce additional copies  
 CC of the polynucleotides. The probes can be used for chromosome mapping of  
 CC the polynucleotide and for detection of transcription levels. Ribozymes  
 CC or antisense oligonucleotides can be generated. The polynucleotides and  
 CC their gene products are used as genetic or biochemical markers (e.g. in  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventive interventions. The polynucleotides, polypeptides and  
 CC antibodies against them can be used in pharmaceutical compositions to

CC treat the cancers and proliferative disorders such as neoplasia,  
 CC dysplasia and hyperplasia.

XX Sequence 360 BP; 123 A; 64 C; 72 G; 101 T; 0 other;

Query Match 61.3%; Score 19; DB 22; Length 360;  
 Best Local Similarity 81.5%; Pred. No. 47;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cagcaggagacagaataaaactgtggaa 28

||| ||||| ||||| ||||| ||||| |||||  
 Db 20 cagaaggagataagaataaaatgtaaaa 46

RESULT 9

AAF65328

ID AAF65328 standard; cDNA; 404 BP.

XX AAF65328;

DT 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 1084.

DE Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
 KW breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a  
 PT mammalian cell and detecting cancer, particularly of the colon or  
 PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 699-700; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to produce probes for detection of mRNA and to produce additional copies  
 CC of the polynucleotides. The probes can be used for chromosome mapping of  
 CC the polynucleotide and for detection of transcription levels. Ribozymes  
 CC or antisense oligonucleotides can be generated. The polynucleotides and  
 CC their gene products are used as genetic or biochemical markers (e.g. in  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventive interventions. The polynucleotides, polypeptides and  
 CC antibodies against them can be used in pharmaceutical compositions to  
 CC treat the cancers and proliferative disorders such as neoplasia,  
 CC dysplasia and hyperplasia.

XX Sequence 404 BP; 135 A; 72 C; 82 G; 112 T; 3 other;

Query Match 61.3%; Score 19; DB 22; Length 404;  
 Best Local Similarity 81.5%; Pred. No. 48;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggacagaataaaactgtgaa 28  
 ||| |||| |||| |||| ||| ||  
 Db 20 cagaaggatagataaaatgtaaaa 46

RESULT 10  
 AAQ76023  
 ID AAQ76023 standard; cDNA; 5578 BP.  
 XX  
 AC AAQ76023;  
 XX  
 DT 10-AUG-1995 (first entry)  
 XX  
 DE TGF-beta-2 promoter region.  
 XX  
 KW TGF-beta-2; transforming growth factor-beta-2; promoter;  
 KW raloxifene responsive element; osteoporosis; therapy; serum lipid;  
 KW hypolipemic; antiestrogen; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT misc\_RNA  
 FT  
 FT Location/Qualifiers  
 1  
 /\*tag= a  
 /note= "base at position 1 corresponds to  
 TGF-beta-2 -227"

TATA\_signal 2248..2252  
 /\*tag= b  
 2278..3980  
 /\*tag= c  
 3981..5578  
 /\*tag= d  
 3635..3980  
 /\*tag= e  
 /codon start= 1

EP629697-A.  
 XX  
 PD 21-DEC-1994.  
 XX  
 PF 20-JUN-1994; 94EP-0304432.  
 XX  
 PR 21-JUN-1993; 93US-0081610.  
 PR 18-MAY-1994; 94US-0246990.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 PI Yang NN;  
 XX  
 DR WPI; 1995-024245/04.  
 XX  
 PT Identification of agents for the treatment of osteoporosis and  
 PT serum lipid lowering agents - using nucleic acids from promoter  
 PT regions of TGF-beta genes comprising regulatory elements  
 PT designated raloxifene responsive elements.  
 XX  
 PS Disclosure; Page 29-32; 65pp; English.  
 XX  
 CC Promoter sequences from TGF-beta, -beta-2 and -beta-3, given in  
 CC AAQ76022-24, respectively, were examined to identify regions  
 CC responsive to the antiestrogen raloxifene. Plasmids encoding a  
 CC reporter gene and partial sequences of the TGF-beta-3 promoter  
 CC identified as raloxifene responsive elements were used to screen  
 CC potential anti-osteoporosis agents.  
 XX  
 SO Sequence 5578 BP; 1498 A; 1324 C; 1257 G; 1499 T; 0 other;

Query Match 61.3%; Score 19; DB 16; Length 5578;  
 Best Local Similarity 81.5%; Pred. No. 76;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 cagggacagaataaaactgtggaat 31  
 |||| |||| |||| |||| |||| ||  
 Db 4316 cagggcagaggttaaaactgaggaatct 4342

RESULT 11  
 AA254436/c  
 ID AA254436 standard; DNA; 476 BP.  
 XX  
 AC AA254436;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria meningitidis ORF 923 partial DNA sequence SEQ ID NO:2821.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO9957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;  
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI; 2000-062150/05.  
 DR P-PSDB; AAY75674.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 7; Page 1329; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254615 to AA254615  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SO Sequence 476 BP; 88 A; 125 C; 118 G; 145 T; 0 other;

Query Match 60.6%; Score 18.8; DB 21; Length 476;

Best Local Similarity 76.7%; Pred. No. 59;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcagcagggacagacagataaaactgtggaata 30  
||||| ||||| ||||| ||||| ||  
Db 445 TCAGCGGGACACAAATAATATAGTGGATTA 416

RESULT 12  
AAZ54437/c  
ID AAZ54437 standard; DNA; 526 BP.  
XX AC AAZ54437;  
XX DT 21-MAR-2000 (first entry)  
XX DE Neisseria meningitidis ORF 923 partial DNA sequence SEQ ID NO:2823.  
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
XX KW antibacterial; gene therapy; ds.  
XX OS Neisseria meningitidis.  
XX PN WO9957280-A2.  
XX PD 11-NOV-1999.  
XX PF 30-APR-1999; 99WO-US09346.  
XX PR 01-MAY-1998; 98US-0083758.  
XX PR 31-JUL-1998; 98US-0094869.  
XX PR 02-SEP-1998; 98US-0098994.  
XX PR 09-SEP-1998; 98US-0099062.  
XX PR 09-OCT-1998; 98US-0103749.  
XX PR 09-OCT-1998; 98US-0103794.  
XX PR 09-OCT-1998; 98US-0103796.  
XX PR 25-FEB-1999; 99US-0121528.  
XX PA (CHIR ) CHIRON CORP.  
XX PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;  
PI Petersen J, Pizzi M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX DR WPI: 2000-062150/05.  
XX DR P-PSDB; AAY75675.  
XX PT Novel Neisserial polypeptides predicted to be useful antigens for  
XX PT vaccines and diagnostics -  
XX PS Claim 7; Page 1330; 1453pp; English.  
XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
XX CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides  
XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
XX CC PCR primers used in the exemplification of the present invention. The  
XX CC polypeptides, the polynucleotides, antibodies and compositions of  
XX CC the invention can be used as vaccines, as diagnostic reagents, and as  
XX CC immunogenic compositions. The polypeptides can be used in the  
XX CC manufacture of medicaments for treating or preventing infection due to  
XX CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
XX CC presence of Neisseria bacteria, or to raise antibodies. They may also  
XX CC be used to screen for agonists or antagonists, which may themselves  
XX CC have use as antibacterial agents. The polynucleotides of the invention  
XX CC may also be used in gene therapy protocols.

XX SQ Sequence 526 BP; 102 A; 135 C; 135 G; 154 T; 0 other;

Query Match 60.6%; Score 18.8; DB 21; Length 526;  
Best Local Similarity 76.7%; Pred. No. 60;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 cagcagggacagacagataaaactgtggaata 31  
||||| ||||| ||||| ||||| ||

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 tcagcagggacagacagataaaactgtggaata 30  
||||| ||||| ||||| ||||| ||  
Db 495 TCAGCGGGACACAAATAATATAGTGGATTA 466

RESULT 13  
AAC57445/c  
ID AAC57445 standard; DNA; 1001 BP.  
XX AC AAC57445;  
XX DT 25-JAN-2001 (first entry)  
XX DE Arachidonic acid metabolism related genomic biallelic marker #79.  
XX KW Human; biallelic marker; arachidonic acid metabolism; genotyping;  
XX KW detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;  
XX KW single nucleotide polymorphism; hybridisation assay; sequencing assay;  
XX KW specific amplification assay; identification; ERBM; 12-LO-RBM;  
XX KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.  
XX OS Homo sapiens.  
XX PN WO200047771-A2.  
XX PD 17-AUG-2000.  
XX PF 11-FEB-2000; 2000WO-IB00184.  
XX PR 12-FEB-1999; 99US-0119917.  
XX PR 23-MAR-1999; 99US-0275267.  
XX PR 07-MAY-1999; 99US-0133200.  
XX PA (GEST ) GENSET.  
XX PI Blumenfeld M, Bougueleret L, Chumakov I;  
XX PI WPI: 2000-571881/53.

PT Novel biallelic markers useful for detecting conditions and genotypes  
PT associated with arachidonic acid metabolism -  
XX PS Claim 13; Page 295; 802pp; English.  
XX CC The present invention describes polynucleotides including biallelic  
XX CC markers derived from genes involved in arachidonic acid metabolism and  
XX CC from genomic regions flanking those genes. Methods from the present  
XX CC invention may be used to select individuals for clinical trials and  
XX CC predict responses to treatment with drugs. The polynucleotides may be  
XX CC used in hybridisation assays, sequencing assays and specific  
XX CC amplification assays for identifying an eicosanoid-related biallelic  
XX CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a  
XX CC segment of nucleotides containing an ERBM. The polynucleotides are  
XX CC useful in diagnostic kits. The markers may be used to detect conditions  
XX CC and genotypes associated with arachidonic acid metabolism. AAC57367 to  
XX CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the  
XX CC exemplification of the present invention.  
XX CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as  
XX CC SNPs) in the polynucleotide sequences from the present invention have  
XX CC been given as their corresponding degenerate bases e.g. a polymorphic  
XX CC base of C or T has been given as Y.  
XX SQ Sequence 1001 BP; 307 A; 202 C; 166 G; 325 T; 1 other;

Query Match 60.6%; Score 18.8; DB 21; Length 1001;  
Best Local Similarity 76.7%; Pred. No. 68;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 cagcagggacagacagataaaactgtggaata 31  
||||| ||||| ||||| ||||| ||

Thu Oct 25 13:08:33 2001

us-09-462-955-1\_copy\_941\_971.rng

Db 596 CAGCAGTGACAGAGTGAACACAGTTTAATGT 567  
RESULT 14  
ID AAC45143  
AC AAC45143 standard; DNA; 1325 BP.  
XX  
AC AAC45143;  
XX  
DT 18-OCT-2000 (first entry)  
DE  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 45452.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.

```

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      60.6%; Score 18.8; DB 21; Length 1325;
Best Local Similarity 76.7%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 cagcaggacagataaaactgtggaatat 31
   | | | | | | | | | | | | | | | | |
Db 1029 cggccggactgaacaaactgtgttat 1058

RESULT 15
ID AAX00614
ID AAX00614 standard; DNA; 2061 BP.
XX

```

```

AC AAX00614;
XX
DT 25-MAR-1999 (first entry)
XX
DE Human secreted protein gene 4 clone HAPBL78.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
XX WO9842738-A1.
XX
XX 01-OCT-1998.
XX
XX 19-MAR-1998; 98WO-0505311.
XX 30-MAY-1997; 97US-0050937.
XX 21-MAR-1997; 97US-0041276.
XX 21-MAR-1997; 97US-0041277.
XX 21-MAR-1997; 97US-0041281.
XX 21-MAR-1997; 97US-0042344.
XX 30-MAY-1997; 97US-0048069.
XX 30-MAY-1997; 97US-0048094.
XX 30-MAY-1997; 97US-0048095.
XX 30-MAY-1997; 97US-0048096.
XX 30-MAY-1997; 97US-0048099.
XX 30-MAY-1997; 97US-0048131.
XX 30-MAY-1997; 97US-0048135.
XX 30-MAY-1997; 97US-0048154.
XX 30-MAY-1997; 97US-0048160.
XX 30-MAY-1997; 97US-0048186.
XX 30-MAY-1997; 97US-0048187.
XX 30-MAY-1997; 97US-0048188.
XX 30-MAY-1997; 97US-0048350.
XX 30-MAY-1997; 97US-0048351.
XX 30-MAY-1997; 97US-0048352.
XX 30-MAY-1997; 97US-0048355.
XX 05-AUG-1997; 97US-0054804.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
PI Rosen CA, Ruben SM, Shi Y, Young P;
XX
XX WPI: 1999-070066/06.
XX P-PSDB: AAW67810.
XX
XX New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1: Page 170-171; 385pp; English.
XX
XX This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived
CC from, are detailed in the descriptor line. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AAX00602) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic
CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in

```

Thu Oct 25 13:08:33 2001

CC the new polynucleotides. Specific uses are described for each of the 87  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAX00611 for described uses).

XX  
 SQ Sequence 2061 BP; 665 A; 414 C; 494 G; 485 T; 3 other;

Query Match 60.6%; Score 18.8; DB 20; Length 2061;  
 Best Local Similarity 76.7%; Pred. No. 77;  
 Matches 23; Conservative: 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcagcaggagacagataaaactgtggaata 30  
 |||| ||| |||| |||| |||| |||| ||||  
 Db 428 tcagaaggacagaaaaatttgtgaaga 457

Search completed: October 24, 2001, 10:07:11  
 Job time: 616 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:35 ; Search time 5479.82 Seconds  
(without alignments)  
53.476 Million cell updates/sec

Title: US-09-462-955-1-copy\_941\_971

Perfect score: 31

Sequence: 1 tcagcaggacagacagataaaactgtggaatat 31

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_estl:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est25:\*
- 26: gb\_est26:\*
- 27: gb\_est27:\*
- 28: gb\_est28:\*
- 29: gb\_est29:\*
- 30: gb\_est30:\*
- 31: gb\_est31:\*
- 32: gb\_est32:\*
- 33: gb\_est33:\*
- 34: gb\_est34:\*
- 35: gb\_est35:\*
- 36: gb\_est36:\*
- 37: gb\_est37:\*
- 38: gb\_est38:\*
- 39: gb\_est39:\*
- 40: gb\_est40:\*
- 41: gb\_est41:\*
- 42: gb\_est42:\*
- 43: gb\_est43:\*
- 44: gb\_est44:\*
- 45: gb\_est45:\*
- 46: gb\_est46:\*
- 47: gb\_est47:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.





Tel: 301 504 8410  
Fax: 301 504 8414

Library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."

BASE COUNT 131 a 110 c 106 g 163 t  
ORIGIN

Query Match .67.7%; Score 21; DB 136; Length 510;  
Best Local Similarity .82.8%; Pred. No. 1.9e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggacagagaataaaactgtggaata 30  
||||| ||||||| ||||||| |||  
Db 150 CACGAGGACAGAAATATCACTGTCAATA 122

RESULT 5  
AZ383240 587 bp DNA GSS 02-OCT-2000  
LOCUS  
DEFINITION IM0140N13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0140N13 R, DNA sequence.  
ACCESSION AZ383240  
VERSION  
KEYWORDS  
SOURCE GSS.  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 587)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0140 row: N column: 13  
Seq primer: CACAGGAAACAGCTATGAC  
Class: plasmid ends  
High quality sequence stop: 587.  
Location/Qualifiers  
1..587  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0140N13"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and

# FEATURES

source  
1..587  
Location/Qualifiers  
1..222  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2306L9"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"  
BASE COUNT 67 a 52 c 34 g 69 t  
ORIGIN

Query Match .67.1%; Score 20.8; DB 223; Length 222;  
Best Local Similarity .91.7%; Pred. No. 2e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cagcaggacagagaataaaactgtg 25  
|| ||||| ||||||| |||||  
Db 219 CACCAGGAAAGAAATAAACTGTG 196

RESULT 7

BASE COUNT 175 a 103 c 114 g 194 t  
ORIGIN

Query Match .67.7%; Score 21; DB 242; Length 587;  
Best Local Similarity .82.8%; Pred. No. 1.9e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tcagcaggacagagaataaaactgtggaat 29  
||||| ||||||| ||||||| |||  
Db 504 TCAGAACTGACAGAAATAAACTGTGTATT 532

RESULT 6  
A0018224/c  
LOCUS  
DEFINITION CIT-HSP-2306L9.TR CIT-HSP Homo sapiens genomic clone 2306L9, DNA sequence.  
ACCESSION A0018224  
VERSION  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 222)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.  
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Other GSSs: CIT-HSP-2306L9.TF  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.  
Location/Qualifiers  
1..222  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2306L9"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"  
BASE COUNT 67 a 52 c 34 g 69 t  
ORIGIN

Query Match .67.1%; Score 20.8; DB 223; Length 222;  
Best Local Similarity .91.7%; Pred. No. 2e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cagcaggacagagaataaaactgtg 25  
|| ||||| ||||||| |||||  
Db 219 CACCAGGAAAGAAATAAACTGTG 196

RESULT 7



```

DEFINITION RICS16437A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION D49255
VERSION D49255.1 GI:702964
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE 1 (bases 1 to 416)
AUTHORS Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE Rice cDNA from callus 1995
JOURNAL Unpublished (1995)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp.
Location/Qualifiers
1. .416
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone_lib="Rice green shoot"
/notes="Green shoot (8 days old)"

BASE COUNT 83 a 111 c 139 g 82 t 1 others
ORIGIN

Query Match 66.5%; Score 20.6; DB 156; Length 416;
Best Local Similarity 85.2%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtgga 27
||||| ||||| ||||| |||||
Db 315 TCAGCAGCAGCAGGATAAAACTTTGGA 289

RESULT 10
AU096982/c
LOCUS AU096982 441 bp mRNA EST 30-JUN-2000
DEFINITION AU096982 Rice green shoot Oryza sativa cDNA clone SI6790, mRNA
sequence.
ACCESSION AU096982
VERSION AU096982.1 GI:8859664
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE 1 (bases 1 to 416)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from green shoot (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.
S16790_1A.
Location/Qualifiers
1. .441
/organism="Oryza sativa"

DEFINITION RICS16437A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION D49255
VERSION D49255.1 GI:702964
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE 1 (bases 1 to 416)
AUTHORS Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE Rice cDNA from callus 1995
JOURNAL Unpublished (1995)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.
S16790_1A.
Location/Qualifiers
1. .416
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="Rice green shoot"
/notes="Green shoot (8 days old)"

BASE COUNT 83 a 111 c 139 g 82 t 1 others
ORIGIN

Query Match 66.5%; Score 20.6; DB 156; Length 416;
Best Local Similarity 85.2%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtgga 27
||||| ||||| ||||| |||||
Db 315 TCAGCAGCAGCAGGATAAAACTTTGGA 289

RESULT 10
AU096982/c
LOCUS AU096982 441 bp mRNA EST 30-JUN-2000
DEFINITION AU096982 Rice green shoot Oryza sativa cDNA clone SI6790, mRNA
sequence.
ACCESSION AU096982
VERSION AU096982.1 GI:8859664
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE 1 (bases 1 to 416)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from green shoot (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.
S16790_1A.
Location/Qualifiers
1. .441
/organism="Oryza sativa"

```

```

/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="SI6790"
/clone_lib="Rice green shoot"
/notes="Green shoot (8 days old)"

BASE COUNT 90 a 120 c 147 g 84 t
ORIGIN

Query Match 66.5%; Score 20.6; DB 107; Length 441;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtgga 27
||||| ||||| ||||| |||||
Db 329 TCAGCAAGCAGCAGGATAAAACTTTGGA 303

RESULT 11
AQ144781/c
LOCUS AQ144781 441 bp DNA GSS 25-SEP-1998
DEFINITION HS_3092_A1_H01_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-3092 Col=1 Row=O, DNA sequence.
ACCESSION AQ144781
VERSION AQ144781.1 GI:3535434
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence tagged Connector
Plate: 3092 row: O column: 1
Class: BAC ends
High quality sequence stop: 441.
Location/Qualifiers
1. .441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3092 Col=1 Row=O"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 107 a 117 c 83 g 134 t
ORIGIN

Query Match 66.5%; Score 20.6; DB 224; Length 441;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 gcaggacagaataaaactgtgga 30
||||| ||||| ||||| |||||
Db 78 GTAGTGTCAAGATAAAACTTTGGAATA 52

RESULT 12
D47654/c

```

```

/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S16437"
/clone_lib="Rice green shoot"
/note="green shoot (8 days old)"
188 a 152 c 143 g 196 t 4 others
BASE COUNT
ORIGIN

Query Match 66.5%; Score 20.6; DB 107; Length 683;
Best Local Similarity 85.2%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcagcagggcagagaataaaaactgtgga 27
110 TCAGCAAGCACAGGATAAAACTTTGGA 84

```

|            |   |                               |                   |             |
|------------|---|-------------------------------|-------------------|-------------|
| RESULT 14  | AW255519/c  | 373 bp                        | EST               | 23-AUG-2000 |
| LOCUS      | MI546   | peppermint glandular trichome | Mentha x piperita | CDNA, mRNA  |
| DEFINITION | sequence.   |                               |                   |             |
| ACCESSION  | AW255519  |                               |                   |             |
| VERSION    | AW255519.1  | GI:7244771                    |                   |             |
| KEYWORDS   | EST.  |                               |                   |             |
| SOURCE     | peppermint.   |                               |                   |             |
| ORGANISM   | Mentha x piperita   |                               |                   |             |
|            | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  |                               |                   |             |
|            | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  |                               |                   |             |
|            | Asteridae; euasterids I; Lamiales; Lamiaceae; Mentha.   |                               |                   |             |
| REFERENCE  | 1 (bases 1 to 373)  |                               |                   |             |
| AUTHORS    | Lange,B.M., Wildung,M.R., Stauber,E.J., Sanchez,C., Pouchnik,D. and<br>Croteau,R.   |                               |                   |             |
| TITLE      | Probing essential oil biosynthesis and secretion by functional<br>evaluation of expressed sequence tags from mint glandular trichomes |                               |                   |             |
| JOURNAL    | proc. Natl. Acad. Sci. U.S.A. 97 (6), 2934-2939 (2000)  |                               |                   |             |
| MEDLINE    | 20183992  |                               |                   |             |
| COMMENT    | Contact: Lange, B.M.<br>Institute of Biological Chemistry/Washington State University<br>Pullman, WA<br>Email: lange-m@mail.wsu.edu.  |                               |                   |             |

```

FEATURES
source
    1. 373
    Location/Qualifiers
        /organism="Mentha x piperita"
        /cultivar="Black Mitcham"
        /db_xref="taxon:34256"
        /clone_lib="peppermint glandular trichome"
        /tissue_type="pellate glandular trichomes"
        /cell_type="secretory"
        /note="Vector: lambda ZAPII"
BASE COUNT      107 a      80 c      81 g      105 t
ORIGIN
Query Match      65.8%;      Score 20.4;      DB 113;      Length 373;
Best Local Similarity 80.0%;      Pred. No. 3.1e+02;
0: Gaps          0: Gaps          6: Indels          0: Gaps

```

|  |     |                                      |
|--|-----|--------------------------------------|
| Qy   | 2   | cagcagggcacagaataaaaactgtggaatat<br> |
| Db   | 237 | CAGGAGGTGAATCAAACTGTGGGATAT<br>      |
| <br>   |     |                                      |
| RESULT 15  |     |                                      |
| BF515120   |     |                                      |
| LOCUS BF515120 425 bp mRNA   |     |                                      |
| DEFINITION UI-H-BMI-ant-e-09-0-UI s1 NCI CGAP_Sub7 Homo sapiens cDNA clone |     |                                      |
| IMAGE:3083632 3', mRNA sequence.   |     |                                      |
| <br>   |     |                                      |
| ACCESSION BF515120   |     |                                      |
| VERSION BF515120.1 GI:11600299   |     |                                      |
| KEYWORDS EST.  |     |                                      |

|            |   |                 |
|------------|---|-----------------|
| RESULT     | 13  |                 |
| AU089813/c |   |                 |
| LOCUS      | AU089813  | 683 bp mRNA     |
| DEFINITION | AU089813 Rice green shoot Oryza sativa cDNA clone S16437, mRNA  | EST 27-APR-2000 |
| ACCESSION  | AU089813  | sequence.       |
| VERSION    | AU089813.1  | GI:7652248      |
| KEYWORDS   | EST.  |                 |
| SOURCE     | Oryza sativa.   |                 |
| ORGANISM   | Oryza sativa  |                 |
|            | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  |                 |
| REFERENCE  | 1 (bases 1 to 683)  |                 |
| AUTHORS    | Sasaki, T. and Yamamoto, K.   |                 |
| TITLE      | Rice cDNA from green shoot (2000)   |                 |
| JOURNAL    | Unpublished (2000)  |                 |
| COMMENT    | Contact: Takuji Sasaki<br>National Institute of Agrobiological Resources<br>Rice Genome Research Program<br>2-1-2 Kannondai, Tsukuba<br>Ibaraki,<br>Japan 305<br>Tel: 0298-38-7441<br>Fax: 0298-38-7468<br>Email: tsasaki@abr.affrc.go.jp<br>PROJECT = 'RGP'.<br>S16437_87. |                 |
| FEATURES   | Location/Qualifiers   |                 |
|            | 1..683  |                 |
|            | source  |                 |

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 425)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. cDNA library preparation: M.B. Soares Lab Clone Distribution:  
 NCI-CGAP clone distribution information can be found through the  
 I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
 Seq primer: M13 Forward  
 POLYA=yes.

FEATURES  
 source

Location/Qualifiers  
 1..425  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3083632"  
 /clone\_lib="NCI-CGAP\_Sub7"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT7T3D-pac (Pharmacia) with a modified  
 polylinker. Site 1: Not 1; Site 2: Eco RI; NCI-CGAP\_Sub7  
 is a subtracted library derived from NCI-CGAP\_Sub6. The  
 NCI-CGAP\_Sub7 library had 12 million recombinants. A  
 single-stranded DNA preparation of NCI-CGAP\_Sub6 was used  
 as a tracer in a subtractive hybridization with a driver  
 comprising: the IMAGE pool (NCI-CGAP\_Kid3 pool 1 LLAM  
 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)  
 1322376-1323911, 1456008-1456775, 1500552-1502855);  
 NCI-CGAP\_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778  
 (IMAGE Clones); NCI-CGAP\_Lu5 pool 1 LLAM 3575-3582,  
 1492104-1493255); NCI-CGAP\_Lu5 pool 1 LLAM 3575-3582,  
 3851-3854 (IMAGE Clones); 1414920-1417991, 1520904-1522439  
 ); NCI-CGAP\_GC4 pool 1 LLAM 3164-3167, 3716-3720,  
 3733-3735 (IMAGE Clones); 1257096-1258631, 1469064-1470983  
 , 1475592-1476743); NCI-CGAP\_Pr22 pool 1 LLAM 2457-2459,  
 2758-2759, 3062-3068 (IMAGE Clones); 985608-986759,  
 1101192-1101959, 1217928-1220615); NCI-CGAP\_Col10 pool 1  
 LLAM 2644-2653, 2871-2872 (IMAGE Clones); 1057416-1061255  
 , 1144584-1145351). (6% of the driver population), plus a  
 pool of 3,840 arrayed clones from NCI-CGAP\_Sub1 (IMAGE  
 Clones 2708616-2710535) and NCI-CGAP\_Sub2 (IMAGE  
 Clones 2710536-2712455) (4% of the driver population  
 ), plus a pool of 11,136 clones from NCI-CGAP\_Sub3 (IMAGE  
 Clones 2712456-2723591) (10% of the driver population),  
 plus a pool of 5,472 clones from NCI-CGAP\_Sub4 (IMAGE  
 Clones 2723592-2729326) (40% of the driver population),  
 plus a pool of 4032 clones from NCI-CGAP\_Sub6 (IMAGE  
 Clones 2728969-2733190) (40% of the driver population).  
 Subtraction was performed as previously described [Bonald  
 , Lennon & Soares (1996): Normalization and Subtraction:  
 Two Approaches to Facilitate Gene Discovery. Genome  
 Research 6, 791-806.  
 TAG\_LIB=NCI-CGAP\_Ov18  
 TAG\_TISSUE=ovary  
 TAG\_SEQ=GCACA"

BASE COUNT 127 a 91 g 137 t  
 ORIGIN

Query Match 65.8%; Score 20.4; DB 149; Length 425;  
 Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 tcacgaggacagataaaactgtgaata 30  
 ||||| ||| | |||| | ||||| |||||

Db 61 TCAGTAGGCAAGAAAAACACTGTGTAATA 90  
 Search completed: October 24, 2001, 13:13:38  
 Job time: 11803 sec

us-09-462-955-1\_copy\_941\_971.rst

Thu Oct 25 13:08:36 2001



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: October 24, 2001, 10:00:12 ; Search time 180.6 Seconds  
(without alignments)  
32.495 Million cell updates/sec

Title: US-09-462-955-1-copy\_941\_971

Perfect score: 31

Sequence: 1 tcagcaggacagataaaactgtggaatat 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTGUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 19    | 61.3        | 5578   | 1     | US-08-081-610-2   |
| 2          | 18.8  | 60.6        | 1957   | 4     | US-09-352-990-11  |
| 3          | 18.8  | 60.6        | 2376   | 3     | US-08-537-361E-3  |
| 4          | 18.8  | 60.6        | 2379   | 3     | US-08-537-361E-5  |
| 5          | 18.8  | 60.6        | 2318   | 1     | US-08-326-670A-1  |
| 6          | 18.8  | 60.6        | 3318   | 3     | US-08-990-470A-1  |
| 7          | 18.8  | 60.6        | 3319   | 3     | US-08-537-361E-1  |
| 8          | 18.2  | 58.7        | 1940   | 1     | US-08-429-742-3   |
| 9          | 18.2  | 58.7        | 4550   | 4     | US-09-103-663-35  |
| 10         | 18.2  | 58.1        | 87350  | 3     | US-08-781-891-79  |
| 11         | 17.6  | 56.8        | 2653   | 1     | US-08-325-553-1   |
| 12         | 17.6  | 56.8        | 2653   | 2     | US-08-394-152A-1  |
| 13         | 17.6  | 56.8        | 3879   | 1     | US-08-530-010-1   |
| 14         | 17.6  | 56.8        | 3879   | 1     | US-08-484-101B-1  |
| 15         | 17.2  | 55.5        | 1094   | 4     | US-09-173-914-5   |
| 16         | 17.2  | 55.5        | 1500   | 1     | US-08-117-083-67  |
| 17         | 17.2  | 55.5        | 4195   | 1     | US-08-340-011-1   |
| 18         | 17.2  | 55.5        | 4195   | 3     | US-08-901-710-1   |
| 19         | 17.2  | 55.5        | 4416   | 3     | US-08-795-430-1   |
| 20         | 17.2  | 55.5        | 4425   | 1     | US-08-222-616-31  |
| 21         | 17.2  | 55.5        | 4425   | 5     | PCT-US95-04228-31 |
| 22         | 17.2  | 55.5        | 4795   | 1     | US-08-340-011-3   |
| 23         | 17.2  | 55.5        | 4795   | 3     | US-08-901-710-3   |
| 24         | 17.2  | 55.5        | 9108   | 5     | PCT-US95-04228-45 |
| 25         | 17    | 54.8        | 2073   | 4     | US-09-032-523-6   |
| 26         | 17    | 54.8        | 2406   | 1     | US-08-586-272-1   |
| 27         | 17    | 54.8        | 2406   | 3     | US-09-082-969-1   |

|      |      |      |       |   |                   |                   |
|------|------|------|-------|---|-------------------|-------------------|
| 28   | 17   | 54.8 | 2498  | 4 | US-09-041-236-1   | Sequence 1, Appli |
| c 29 | 17   | 54.8 | 13417 | 2 | US-08-637-759B-37 | Sequence 37, Appl |
| c 30 | 17   | 54.8 | 13417 | 3 | US-08-871-355A-37 | Sequence 37, Appl |
| c 31 | 17   | 54.8 | 51952 | 3 | US-08-947-823-1   | Sequence 1, Appli |
| c 32 | 16.8 | 54.2 | 366   | 2 | US-08-672-345C-87 | Sequence 87, Appl |
| c 33 | 16.8 | 54.2 | 366   | 2 | US-08-672-345C-89 | Sequence 89, Appl |
| c 34 | 16.8 | 54.2 | 720   | 2 | US-07-956-399-1   | Sequence 1, Appli |
| c 35 | 16.8 | 54.2 | 1716  | 3 | US-08-656-034-9   | Sequence 9, Appli |
| c 36 | 16.8 | 54.2 | 1816  | 2 | US-08-951-148-2   | Sequence 2, Appli |
| c 37 | 16.8 | 54.2 | 1816  | 2 | US-09-165-234-2   | Sequence 2, Appli |
| c 38 | 16.8 | 54.2 | 1816  | 3 | US-09-274-570-2   | Sequence 2, Appli |
| c 39 | 16.8 | 54.2 | 2160  | 3 | US-08-656-034-1   | Sequence 2, Appli |
| c 40 | 16.8 | 54.2 | 2562  | 1 | US-08-146-421-4   | Sequence 4, Appli |
| c 41 | 16.6 | 53.5 | 264   | 4 | US-08-991-789A-24 | Sequence 24, Appl |
| c 42 | 16.6 | 53.5 | 1240  | 2 | US-08-874-186-39  | Sequence 39, Appl |
| c 43 | 16.6 | 53.5 | 2240  | 2 | US-08-671-978A-5  | Sequence 5, Appli |
| c 44 | 16.6 | 53.5 | 2953  | 2 | US-08-859-201-1   | Sequence 1, Appli |
| c 45 | 16.6 | 53.5 | 3286  | 4 | US-09-211-417-2   | Sequence 2, Appli |

## ALIGNMENTS

RESULT 1  
US-08-081-610-2  
; Sequence 2, Application US/08081610  
; Patent No. 5445941  
; GENERAL INFORMATION:  
; APPLICANT: Yang, Na N  
; TITLE OF INVENTION: Materials and Methods for Screening  
; TITLE OF INVENTION: Anti-Osteoporosis or Serum Lipid Lowering Agents  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti and Witcoff, Ltd.  
; STREET: 10 S. Wacker Dr.  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: U.S.A  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/081.610  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heaphy, Barbara A  
; REGISTRATION NUMBER: 34,619  
; REFERENCE/DOCKET NUMBER: 93,402  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5578 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: TATA\_signal  
; LOCATION: 2248..2252  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 2278..3980  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 3981..5578  
; FEATURE:  
; NAME/KEY: misc\_rna

Thu Oct 25 13:08:35 2001

LOCATION: 3635..3980  
OTHER INFORMATION: /note= "CDS, Codon start = 1"  
FEATURE:  
NAME/KEY: misc\_RNA  
LOCATION: 1..2  
OTHER INFORMATION: /note= "Number 1 corresponds to  
OTHER INFORMATION: TGF-2 -2277"  
US-08-081-610-2

Query Match 61.3%; Score 19; DB 1; Length 5578;  
Best Local Similarity 81.5%; Pred. No. 14; 5; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 5 caggcagacagataaaactgtggaatat 31  
||||| ||||| ||||| ||||| |||||

Db 4316 CAGGGCAGAGTAACTGAGGAATCT 4342

## RESULT 2

US-09-352-990-11  
Sequence 11, Application US/09352990

Patent No. 6255090

GENERAL INFORMATION:

APPLICANT: Famodu, Layo O.

APPLICANT: Orozco, Buddy

APPLICANT: Rafalski, Antoni

TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase

FILE REFERENCE: BB-1191

CURRENT APPLICATION NUMBER: US/09/352,990

CURRENT FILING DATE: 1999-07-14

EARLIER APPLICATION NUMBER: 60/092,866

EARLIER FILING DATE: July 15, 1998

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Microsoft Office 97

SEQ ID NO 11

LENGTH: 1957

TYPE: DNA

ORGANISM: Oryza sativa

US-09-352-990-11

Query Match 60.6%; Score 18.8; DB 4; Length 1957;  
Best Local Similarity 76.7%; Pred. No. 14;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 cagcaggacagataaaactgtggaatat 31  
||||| ||||| ||||| ||||| |||||

Db 1323 cagcaggacagataaaactgtggaatat 1352

## RESULT 3

US-08-537-361E-3

Sequence 3, Application US/08537361E

Patent No. 6121037

GENERAL INFORMATION:

APPLICANT: Stojiljkovic, Igor

APPLICANT: So, Magdalene

APPLICANT: Hwa, Vivian

APPLICANT: Heffron, Fred

APPLICANT: Nassif, Xavier

TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor

TITLE OF INVENTION: Genes and Uses

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

STREET: 300 South Wacker Drive, 32nd Floor

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/537,361E

FILING DATE: 02-OCT-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: No. 6121037han, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 94,784-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-0001

TELEFAX: 312-913-0002

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2376 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2373

US-08-537-361E-3

Query Match 60.6%; Score 18.8; DB 3; Length 2376;  
Best Local Similarity 76.7%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcagcaggacagataaaactgtggaata 30  
||||| ||||| ||||| ||||| |||||

Db 828 TCAGCAGGGCAGTAACTACAGGTTGAAGA 857

## RESULT 4

US-08-537-361E-5

Sequence 5, Application US/08537361E

Patent No. 6121037

GENERAL INFORMATION:

APPLICANT: Stojiljkovic, Igor

APPLICANT: So, Magdalene

APPLICANT: Hwa, Vivian

APPLICANT: Heffron, Fred

APPLICANT: Nassif, Xavier

TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor

TITLE OF INVENTION: Genes and Uses

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

STREET: 300 South Wacker Drive, 32nd Floor

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/537,361E

FILING DATE: 02-OCT-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: No. 6121037han, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 94,784-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-0001

TELEFAX: 312-913-0002

INFORMATION FOR SEQ ID NO: 5:

```

; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 2379 base pairs
; ; TYPE: nucleic acid.
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: DNA (genomic)
; ; FEATURE:
; ; NAME/KEY: CDS
; ; LOCATION: 1..2376
US-08-537-361E-5

```

|                       |              |               |               |              |
|-----------------------|--------------|---------------|---------------|--------------|
| Query Match           | 60.6%;       | Score 18.8;   | DB 3;         | Length 2379; |
| Best Local Similarity | 76.7%;       | Pred. No. 15; |               |              |
| Matches 23;           | Conservative | 0;            | Mismatches 7; | Indels 0     |

Oy 1 tcagcagggacagaataaaactgtggaata 30  
 |||||  
 Db 828 TCAGCAGGGGCATATTTACACGGTTGAAGA 857

RESULT 5  
US-08-326-670A-1  
; Sequence 1, Application US/08326670A  
; Patent No. 5698438

Query Match 60.6%; Score 18.8; DB 1; Length 3318;  
Best Local Similarity 76.7%; Pred. No. 16;  
Matches 23; Conservative 0; Mismatches 7; Indels 0

Qy 1 tcagcagggacagaataaaaactgtggaata 30  
|||||  
Db 1297 TCAGCAGGGGCATATTACACGGTTGAAGA 1326

```

RESULT      6
US-08-990-470A-1
; Sequence 1, Application US/08990470A
; Patent No. 6123942
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6123942a1 Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:

```

Query Match 60.6%; Score 18.8; DB 3; Length 3318;  
Best Local Similarity 76.7%; Pred. No. 16;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcagcagggacagaataaaaactgtggaata 30  
 |||||  
 Db 1297 TCAGCAGGGGCATAATTACACGGTTGAAGA 1326

```

RESULT      7
US-0537-351E-1
; Sequence 1, Application US/08537351E
; Patent No. 6121037
;
; GENERAL INFORMATION:
;
; APPLICANT: Stojiljkovic, Igor
;
; APPLICANT: So, Magdalena
;
; APPLICANT: Iwa, Vivian
;

```

```

;
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,361E
; FILING DATE: 02-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6121037nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 471..2848
;
; US-08-537-361E-1
;
;
; Query Match 60.6%; Score 18.8; DB 3; Length 3319;
; Best Local Similarity 76.7%; Pred. No. 16;
; Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
;
; Qy 1 tcagcagggacagaataaactgtggaata 30
; ||||| ||||| ||||| ||||| |||||
; Db 1298 TCAGCAGGGGCATAATTACACGGTTCAAGA 1327
;
;
; RESULT 8
; US-08-429-742-3
; Sequence 3, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,742
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0505
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1940 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..1197
;
; US-08-429-742-3
;
;
; Query Match 58.7%; Score 18.2; DB 1; Length 1940;
; Best Local Similarity 87.0%; Pred. No. 26;
; Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; Qy 3 agcagggacagaataaactgtg 25
; ||||| ||||| ||||| |||||
; Db 1797 AGTAGGACAGACAGCAAACTGTG 1819
;
;
; RESULT 9
; US-09-103-663-35/C
; Sequence 35, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:
; APPLICANT: Kinet et al.
; TITLE OF INVENTION: Isolation, characterization, and use of the human beta
; TITLE OF INVENTION: subunit of the high affinity receptor for
; TITLE OF INVENTION: immunoglobulin E.
; FILE REFERENCE: 50490
; CURRENT APPLICATION NUMBER: US/09/103,663D
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 07/869,933
; EARLIER FILING DATE: 1992-04-16
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 4550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (578)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (735)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1362)
; OTHER INFORMATION: n represents a, c, t or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2479)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2517)
; OTHER INFORMATION: n represents a, c, t or g.
;
;
; US-09-103-663-35/C
```

```

;
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,361E
; FILING DATE: 02-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6121037nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 471..2848
;
; US-08-537-361E-1
;
;
; Query Match 60.6%; Score 18.8; DB 3; Length 3319;
; Best Local Similarity 76.7%; Pred. No. 16;
; Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
;
; Qy 1 tcagcagggacagaataaactgtggaata 30
; ||||| ||||| ||||| ||||| |||||
; Db 1298 TCAGCAGGGGCATAATTACACGGTTCAAGA 1327
;
;
; RESULT 8
; US-08-429-742-3
; Sequence 3, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,742
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0505
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1940 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..1197
;
; US-08-429-742-3
;
;
; Query Match 58.7%; Score 18.2; DB 1; Length 1940;
; Best Local Similarity 87.0%; Pred. No. 26;
; Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; Qy 3 agcagggacagaataaactgtg 25
; ||||| ||||| ||||| |||||
; Db 1797 AGTAGGACAGACAGCAAACTGTG 1819
;
;
; RESULT 9
; US-09-103-663-35/C
; Sequence 35, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:
; APPLICANT: Kinet et al.
; TITLE OF INVENTION: Isolation, characterization, and use of the human beta
; TITLE OF INVENTION: subunit of the high affinity receptor for
; TITLE OF INVENTION: immunoglobulin E.
; FILE REFERENCE: 50490
; CURRENT APPLICATION NUMBER: US/09/103,663D
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 07/869,933
; EARLIER FILING DATE: 1992-04-16
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 4550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (578)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (735)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1362)
; OTHER INFORMATION: n represents a, c, t or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2479)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2517)
; OTHER INFORMATION: n represents a, c, t or g.
;
;
; US-09-103-663-35/C
```

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2526)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2549)..(2552)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2612)..(2614)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2633)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2920)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3069)  
OTHER INFORMATION: n represents a, c, t or g.  
US-09-103-663-35

Query Match 58.7%; Score 18.2; DB 4; Length 4550;  
Best Local Similarity 74.2%; Pred. No. 31;  
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 tcagcaggacagaataaaactgtggaatat 31  
Db 4472 TTACGACAGACTGAATAAATCTATGATAAAT 4442

RESULT 10  
US-08-781-891-79/c  
Sequence 79, Application US/08781891  
Patent No. 6090620  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-En  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781.891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620/enburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:  
LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-79

Query Match 58.1%; Score 18; DB 3; Length 87350;  
Best Local Similarity 80.8%; Pred. No. 68;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 agggacagagaataaaactgtggaatat 31  
Db 49985 AGGTACAGCAATAAAACTGTAGCCCTTT 49960

RESULT 11  
US-08-325-553-1  
Sequence 1, Application US/08325553  
Patent No. 5538866  
GENERAL INFORMATION:  
APPLICANT: Israeli, Ron S.  
APPLICANT: Heston, Warren D.W.  
APPLICANT: Fair, William R.  
TITLE OF INVENTION: THE PROSTATE-SPECIFIC MEMBRANE ANTIGEN  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/325.553  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/973.337A  
FILING DATE: 05 NOV 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/41426  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2653 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Carcinoma  
IMMEDIATE SOURCE:  
CLONE: Prostate-Specific Membrane Antigen  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 262..2511  
US-08-325-553-1

```

Query Match      56.8%; Score 17.6; DB 1; Length 2653;
Best Local Similarity 83.3%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggagacagaataaaactgt 24
   | ||||| ||||| ||||| |||||
DB 508 TTAGCAGGACAGACAGAACTTT 531

RESULT 12
US-08-394-152A-1
; Sequence 1, Application US/08394152A
; Patent No. 5935818
; GENERAL INFORMATION:
; APPLICANT: Israeli, Ron S.
; APPLICANT: Heston, Warren D.W.
; APPLICANT: Fair, William R.
; TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,152A
; FILING DATE: 24-FEB-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41426-B
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2653 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Carcinoma
; IMMEDIATE SOURCE:
; CLONE: Prostate-Specific Membrane Antigen
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262..2511
US-08-394-152A-1

Query Match      56.8%; Score 17.6; DB 2; Length 2653;
Best Local Similarity 83.3%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggagacagaataaaactgt 24
   | ||||| ||||| ||||| |||||
DB 508 TTAGCAGGACAGACAGAACTTT 531

us-09-462-955-1_copy_941_971.rni

RESULT 13
US-08-530-010-1/c
; Sequence 1, Application US/08530010
; Patent No. 5689055
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliott M.
; APPLICANT: Chang, Caren
; APPLICANT: Bleeker, Anthony B.
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,010
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/086,555
; FILING DATE: 01-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-57515/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 998-3249
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-530-010-1

Query Match      56.8%; Score 17.6; DB 1; Length 3879;
Best Local Similarity 83.3%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 caggagacagaataaaactgtgaa 28
   | ||||| ||||| ||||| |||||
DB 455 CAACAACAGAGAAACTGTGAA 432

RESULT 14
US-08-484-101B-1/c
; Sequence 1, Application US/08484101B
; Patent No. 5824868
; GENERAL INFORMATION:
; APPLICANT: California Institute of Technology
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
; TITLE OF INVENTION: ETHYLENE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

Search completed: October 24, 2001, 10:00:14  
Job time: 199 sec

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,101B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/  
FILING DATE: 01-JUL-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/086,555  
FILING DATE: 01-JUL-1993  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-57515-2/RFT  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-484-101B-1

Query Match 56.8%; Score 17.6; DB 1; Length 3879;  
Best Local Similarity 83.3%; Pred. No. 55;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 5 cagggacagaaataaaactgtggaa 28  
||| ||||| ||||| ||||| |||||  
Db 455 CACACAGAGAGAAACTGTGGAA 432  
  
RESULT 15  
US-09-173-914-5/c  
Sequence 5, Application US/09173914  
Patent No. 6171857  
GENERAL INFORMATION:  
APPLICANT: Hendrickson, Eric  
TITLE OF INVENTION: A No. 6171857el Leucine Zipper, KARP-1 and  
METHOD OF INVENTION: Methods of Regulating DNA Dependent Protein Kinase Activity  
FILE REFERENCE: B0877/7017/HK  
CURRENT APPLICATION NUMBER: US/09/173,914  
CURRENT FILING DATE: 1998-10-16  
EARLIER APPLICATION NUMBER: 60/064,557  
EARLIER FILING DATE: 1997-10-17  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows version 3.0  
SEQ ID NO 5  
LENGTH: 1094  
TYPE: DNA  
ORGANISM: Salmonella Typhimurium  
US-09-173-914-5

Query Match 55.5%; Score 17.2; DB 4; Length 1094;  
Best Local Similarity 73.3%; Pred. No. 64;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
  
QY 1 tcagcaggacagaaataaaactgtggaata 30  
||||| ||||| ||||| ||||| |||||  
Db 1074 TCAGCAGGACAGACGTGTGTTGATTA 1045

us-09-462-955-1\_copy\_941\_971.rni

Thu Oct 25 13:08:35 2001



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:11 ; Search time 393.52 seconds  
(without alignments)  
673.344 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_583\_1004  
Perfect score: 422  
Sequence: 1 cttgaagtgtgtgtgcgat.....catctaaattacacccaatac 422

Scoring table: IDENTITY\_NUC  
Gapop 10.0'', Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn/NA1980.DAT.\*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn/NA1981.DAT.\*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn/NA1982.DAT.\*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn/NA1983.DAT.\*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn/NA1984.DAT.\*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn/NA1985.DAT.\*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn/NA1986.DAT.\*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn/NA1987.DAT.\*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn/NA1988.DAT.\*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn/NA1989.DAT.\*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn/NA1990.DAT.\*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn/NA1991.DAT.\*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn/NA1992.DAT.\*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn/NA1993.DAT.\*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn/NA1994.DAT.\*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn/NA1995.DAT.\*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn/NA1996.DAT.\*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn/NA1997.DAT.\*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn/NA1998.DAT.\*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn/NA1999.DAT.\*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn/NA2000.DAT.\*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID          | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 114.2 | 27.1        | 1096   | 20 AAV71833 | BBTV DNA II clone  |
| 2          | 114.2 | 27.1        | 1096   | 21 AAA38950 | Banana bunchy top  |
| 3          | 112.6 | 26.7        | 1091   | 20 AAV71834 | BBTV DNA II clone  |
| 4          | 112.6 | 26.7        | 1091   | 21 AAA38947 | Banana bunchy top  |
| 5          | 112.4 | 26.6        | 886    | 20 AAA34687 | DNA sequence of BB |
| 6          | 96.8  | 22.9        | 1106   | 20 AAV71832 | BBTV DNA I clone ( |
| 7          | 95.4  | 22.6        | 1017   | 17 AAT13165 | SCSV segment 6. S  |
| 8          | 95.2  | 22.6        | 1106   | 21 AAA38946 | Banana bunchy top  |
| 9          | 88.4  | 20.9        | 1022   | 17 AAT13161 | SCSV segment 2. S  |
| 10         | 80.4  | 19.1        | 593    | 20 AAA34686 | DNA sequence of BB |
| 11         | 62.8  | 14.9        | 1110   | 19 AAV24089 | Banana bunchy top  |

|    |      |      |       |             |                    |
|----|------|------|-------|-------------|--------------------|
| 12 | 59.6 | 14.1 | 982   | 18 AAT49405 | Banana bunchy top  |
| 13 | 59.6 | 14.1 | 1110  | 19 AAV24086 | Banana bunchy top  |
| 14 | 59.6 | 14.1 | 1111  | 19 AAV24077 | Banana bunchy top  |
| 15 | 59.6 | 14.1 | 1111  | 19 AAV24084 | Banana bunchy top  |
| 16 | 59.6 | 14.1 | 1111  | 19 AAV24087 | Banana bunchy top  |
| 17 | 58.6 | 13.9 | 1103  | 19 AAV24091 | Banana bunchy top  |
| 18 | 58.6 | 13.9 | 1105  | 19 AAV24093 | Banana bunchy top  |
| 19 | 58   | 13.7 | 1109  | 19 AAV24088 | Banana bunchy top  |
| 20 | 58   | 13.7 | 1111  | 19 AAV24090 | Banana bunchy top  |
| 21 | 57   | 13.5 | 1104  | 19 AAV24092 | Banana bunchy top  |
| 22 | 56.4 | 13.4 | 1111  | 19 AAV24085 | Banana bunchy top  |
| 23 | 49.4 | 11.7 | 936   | 22 AAF58252 | Oligonucleotide D1 |
| 24 | 49.4 | 11.7 | 936   | 22 AAF58254 | Oligonucleotide D1 |
| 25 | 49.4 | 11.7 | 936   | 22 AAF58257 | Oligonucleotide D1 |
| 26 | 49.4 | 11.7 | 936   | 22 AAF58259 | Oligonucleotide D1 |
| 27 | 49.4 | 11.7 | 936   | 22 AAF58262 | Oligonucleotide D2 |
| 28 | 49.4 | 11.7 | 938   | 22 AAF58255 | Oligonucleotide D2 |
| 29 | 46   | 10.9 | 936   | 22 AAF58252 | Oligonucleotide D1 |
| 30 | 46   | 10.9 | 936   | 22 AAF58254 | Oligonucleotide D1 |
| 31 | 46   | 10.9 | 936   | 22 AAF58257 | Oligonucleotide D1 |
| 32 | 46   | 10.9 | 936   | 22 AAF58259 | Oligonucleotide D1 |
| 33 | 46   | 10.9 | 936   | 22 AAF58262 | Oligonucleotide D2 |
| 34 | 46   | 10.9 | 938   | 22 AAF58255 | Oligonucleotide D2 |
| 35 | 42   | 10.0 | 300   | 20 AAV71831 | Subgenomic fragmen |
| 36 | 42   | 10.0 | 300   | 21 AAA38949 | Banana bunchy top  |
| 37 | 41   | 9.7  | 287   | 20 AAV71830 | Subgenomic fragmen |
| 38 | 41   | 9.7  | 287   | 21 AAA38948 | Banana bunchy top  |
| 39 | 38   | 9.0  | 244   | 22 AAF58238 | Oligonucleotide D1 |
| 40 | 38   | 9.0  | 244   | 22 AAF58238 | Oligonucleotide D1 |
| 41 | 32.6 | 7.7  | 5053  | 19 AAV22816 | Nucleotide sequenc |
| 42 | 32.6 | 7.7  | 5055  | 19 AAV17236 | DNA from a region  |
| 43 | 31.4 | 7.4  | 41599 | 21 AAZ35351 | Cosmid including s |
| 44 | 30.8 | 7.3  | 41599 | 21 AAZ35351 | Cosmid including s |
| 45 | 30.6 | 7.3  | 9345  | 21 AAZ93966 | Mouse uromodulin p |

ALIGNMENTS

|          |   |
|----------|---|
| RESULT   | 1   |
| AAV71833 | AAV71833 standard; DNA; 1096 BP.  |
| ID       | AAV71833 standard; DNA; 1096 BP.  |
| XX       |   |
| AC       | AAV71833;   |
| XX       |   |
| DT       | 10-FEB-1999 (first entry)   |
| XX       |   |
| DE       | BBTV DNA II clone (2-17) nucleotide sequence.                             |
| XX       |   |
| KW       | Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;      |
| KW       | Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss. |
| XX       |   |
| OS       | Banana bunchy top virus.  |
| XX       |   |
| FH       | Key   |
| FT       | TATA_signal Location/Qualifiers   |
| FT       | stem_loop 1..7  |
| FT       | /*tag= a  |
| FT       | /*tag= b  |
| FT       | 70..928   |
| FT       | /*tag= c  |
| FT       | /product= "ORF-V2 product"  |
| FT       | /transl_except= (pos:215..217, aa:Gly)                                    |
| FT       | polyA_signal 533..538   |
| FT       | /*tag= d  |
| FT       | polyA_signal 799..804   |
| FT       | /*tag= e  |
| FT       | polyA_signal 907..912   |
| FT       | /*tag= f  |
| FT       | polyA_signal 1030..1035   |
| FT       | /*tag= ge   |
| XX       |   |
| PN       | US5846705-A.  |





|    |     |   |     |
|----|-----|---|-----|
| QY | 19  | atcggagagccagcgacgacgcacaaatcctctggatcgcgacgagacgagagagac   | 78  |
|    |     |   |     |
| Db | 272 | atcgaaggtctcgtgatcgaagtatcatctgggtacggtcccaacgagcgagcga     | 331 |
|    |     |   |     |
| QY | 79  | ggaaagtcctggttggccaaatatcggacctaagcccgactggttctacacatgtgt   | 138 |
|    |     |   |     |
| Db | 332 | ggaaagtcacactcgcaagatatctatcattaaaactggatggggatataccaacggt  | 391 |
|    |     |   |     |
| QY | 139 | ggaaaccagaagaagcattgtaccagtacatcgagaccccaaacgaaatttaactcctc | 198 |
|    |     |   |     |
| Db | 392 | ggaaagacgtcggtatgatgcacatacaacgatggaactgataatactggattatt    | 451 |
|    |     |   |     |
| QY | 199 | gatgtaccacgagtgaatttagagatttaaatatgccctgttagaatgtttaagaac   | 258 |
|    |     |   |     |

CC construction of PCR primers, c

\_\_\_\_\_



```

XX      Sequence 1022 BP; 309 A; 183 C; 245 G; 285 T; 0 other;
SQ
      Query Match      20.9%; Score 88.4; DB 17; Length 1022;
      Best Local Similarity 54.3%; Pred. NO. 1.1e-19;
      Matches 201; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

QY      25  gagccagcgagcagtcgcacaaatcctctggatctcggacagacgagcgagcagcggaag 84
      Db      559  ggggaaccagatgcgagtgatcatctgggctctgtctcagacggaggagagaaag 618
QY      85  tcctgtttgccaaatctcggaatcaagcccgactggttccacatgtggtggaacc 144
      Db      619  acgagcttcgcaaggaaatt---aatcaggtatggtggtttacacgcgaggggaag 675
QY      145  agaaaagagattgtaccagtatcatcgaggaccacaaacgaaatttaatcctcgatga 204
      Db      676  accaggagcgtattatcatgtgatctcaagccacagagagaaatgcgtttgatgtt 735
QY      205  ccaggtgtaatttagatgatttaaatattgcccctgttagaattggttaagaaacaggcca 264
      Db      736  ccaggtgttcttcggagatgatgaactacggcgatggagatgttgaaagacagagtt 795
QY      265  ttcagttcggacaaacacgaaccccttagttatcttgggttcgacatgtgcattactc 324
      Db      796  ttgcgaagtacaaaatataggccctgatagctcttgattaggaagttagttcatttaatt 855
QY      325  gtatttgcgaatgtcctgctgattttgaaaatacagcaggagacagagaataaacctgtgg 384
      Db      856  gtgttgcgaactggcacctgacgtccccacgcgcataagtggaggacagacttgtaattac 915
QY      985  aaattattaa 394
      Db      985  aaattattaa 394

```



Thu Oct 25 13:08:15 2001

Db 287 ggaatttaagaatggaataattcaaaagcgggaaataatgaacccgttttgaagatag--- 343  
 Qy 306 cgaccatgtgcatgtactgtattgtcccaatgtctcgtctgatttgaataatcagcag 365  
 Db 344 agaatagtcgaagtcattgtaattggttaacttctcctccgaaggaagaatcttttctga 403  
 Qy 366 ggacagaataaaactg 381  
 Db 404 agatcgaataaagtgtg 419

## RESULT 13

AAV24086  
ID AAV24086 standard; DNA; 1110 BP.

XX AAV24086;

XX 11-AUG-1998 (first entry)

XX Banana bunchy top virus component 1 DNA sequence.

XX BBTv; probe; diagnostic primer; component 1; ss.

XX Banana bunchy top virus.

XX US5756708-A.

XX 26-MAY-1998.

XX 24-FEB-1994; 94US-0202186.

XX 24-FEB-1994; 94US-0202186.

XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.

XX Burns TM, Dale JL, Harding RM, Karan M;

XX WPI; 1998-321636/28.

XX Isolated Banana bunchy top virus DNA - useful as diagnostic probes  
 PT and primers and for producing virus-resistant plants

XX Disclosure; Column 19-22; 59pp; English.

XX This sequence represents a DNA fragment from the component 1 of  
 CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as  
 CC diagnostic probes or primers or can be inserted into plants or other  
 CC organisms e.g. to produce virus-resistant plants or to act as promoters,  
 CC enhancers or termination signals.

SQ Sequence 1110 BP; 366 A; 191 C; 276 G; 277 T; 0 other;

Query Match 14.1%; Score 59.6; DB 18; Length 982;  
 Best Local Similarity 51.3%; Pred. No. 5.1e-10;  
 Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

Qy 6 agtctctgtcgtatcgagcagcagcggcagcgcacacatctctgtgatatcgcgagc 65  
 Db 53 agaggttgagaataatcgtcgcagcagcagcgcacacatctctgtgatatcgcgagc 112  
 Qy 66 agacggaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 125  
 Db 113 aaatggaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 172  
 Qy 126 ctacacatgtgtggaaccgagcagcagcagcagcagcagcagcagcagcagcagc 185  
 Db 173 ttattctccaggaggaataatcgtcgcagcagcagcagcagcagcagcagcagcagc 229  
 Qy 186 aaatttaactctcgtatcgagcagcagcagcagcagcagcagcagcagcagcagc 245  
 Db 230 ---tgttattgtatctcgaagcagcagcagcagcagcagcagcagcagcagcagc 286  
 Qy 246 atgtgttaagaacaggcagcagcagcagcagcagcagcagcagcagcagcagcagc 305

## RESULT 12

AAAT49405  
ID AAAT49405 standard; DNA; 982 BP.

XX AAAT49405;

XX 23-AUG-1997 (first entry)

XX Banana bunchy top virus DNA intergenic region 1 insert in pBTV1.INT.

XX BBTv; intergenic region; promoter; transgenic plant; ds.

XX Banana bunchy top virus.

XX WO9638554-A1.

XX 05-DEC-1996.

XX 31-MAY-1996; 96WO-AU00335.

XX 31-MAY-1995; 95AU-0003285.

XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.

XX Becker DK, Beetham PR, Dale JL, Dugdale B, Hafner GJ;

XX Harding RM;

XX WPI; 1997-034368/03.

XX DNA from intergenic region of banana bunchy top virus DNA component  
 PT - useful for promoting, enhancing, regulating or modifying  
 PT transcription of a non-BBTv gene in transgenic plants

XX Claim 4; Fig 12; 80pp; English.

XX Intergenic regions (AAAT49399-409) are derived from genomic DNA  
 CC components 1-6 (see also AAAT49386-90) of banana bunchy top virus  
 CC (BBTV). A large intergenic region (AAAT49405) comprises the insert  
 CC in plasmid pBTV1.INT, a plasmid obd. by PCR amplification  
 CC (see also AAAT49427-28) of a full-length BBTv clone and cloning of  
 CC TaqI-digested PCR product into pBl101.3. BBTv intergenic regions  
 CC are useful for promoting, enhancing, regulating or modifying  
 CC transcription of non-BBTv genes in monocotyledon or dicotyledon  
 CC transgenic plants. The non-BBTv gene is e.g. an insecticide  
 CC resistance gene, herbicide resistance gene or growth promoting  
 CC gene.

SQ Sequence 982 BP; 316 A; 174 C; 249 G; 243 T; 0 other;



```

QY 246 atgtgttaagaacagggcattcgttcgacaaatacgaaccccttagtcttgggtt 305
Db 818 ggaatttaagaatgaataattcaaaagcggaataatgaacccgtttttgaagatag---t 874
QY 306 cgaacctgtgcattgactcgtatttggcaatgtcctgcctgctgattatttgaataacagcag 365
Db 875 aqaatatgtcgaagtcattgtaatggctcaacttcttccgaaggaagaaatctttctga 934
QY 366 ggaacagaataaaactg 381
Db 935 agatcgaataaagtgtg 950

RESULT 14
AAV24077
ID AAV24077 standard; DNA; 1111 BP.
XX
AC AAV24077;
XX
DT 11-AUG-1998 (first entry)
XX
DE Banana bunchy top virus component 1 DNA sequence.
XX
KW BBTv; probe; diagnostic primer; component 1; ss.
XX
OS Banana bunchy top virus.
XX
FH Key Location/Qualifiers
FT CDS 129..989
FT /*tag= a
XX
PN US5756708-A.
XX
PD 26-MAY-1998.
XX
PF 24-FEB-1994; 94US-0202186.
XX
PR 24-FEB-1994; 94US-0202186.
XX
PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
XX
PI Burns TM, Dale JL, Harding RM, Karan M;
XX
DR WPI; 1998-321636/28.
XX
P-PSDB; AAW54071.
XX
PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers and for producing virus-resistant plants
XX
PS Example 1; Fig 11; 59pp; English.
XX
CC This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plants or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
CC enhancers or termination signals.
XX
SQ Sequence 1111 BP; 363 A; 192 C; 276 G; 280 T; 0 other;

Query Match 14.1%; Score 59.6; DB 19; Length 1111;
Best Local Similarity 51.3%; Pred. No. 5.4e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 6 agtgcgtctgcgtatcgagagccagcgacgacgcacaaatcctctggatcgagcag 65
Db 611 agaggtggagaataatcattgcgcacccatgctcggagaataatttggctatggccc 670
QY 66 agacggagagagcggagtcggttcttggcaaatatctcggaactcaagcccgactggtt 125
Db 671 aaatggagagaggaagaaacacgctatgcaaaacatctaataagacgagaaatgcgtt 730
QY 126 ctacacatgtgtggaaacgaagagcgtattgtaccgatactacgcagagcccaaacg 185

```

```

Db 731 ttattccaggaggaataatcattggatatactagactgataataatcagagatat--- 787
QY 186 aaatttaattcctgcgtatgccaggtgtaatttagagattatttaaaattatgccctgttaga 245
Db 788 ---tggtatttgatattccaaagatgcaaaagaggattatttaaatattgggttataga 844
QY 246 atgtgttaagaacagggcattcgttcgacaaatacgaaccccttagtcttgggtt 305
Db 845 ggaatttaagaatggaataattcaaaagcggaataatgaacccgttttgaagatag---t 901
QY 306 cgaacctgtgcattgactcgtatttggcaatgtcctgcctgattatttgaataacagcag 365
Db 902 agaatatgtcgaagtcattgtaatggctcaacttcttccgaaggaagaaatctttctga 961
QY 366 ggaacagaataaaactg 381
Db 962 agatcgaataaagtgtg 977

RESULT 15
AAV24084
ID AAV24084 standard; DNA; 1111 BP.
XX
AC AAV24084;
XX
DT 11-AUG-1998 (first entry)
XX
DE Banana bunchy top virus component 1 DNA sequence.
XX
KW BBTv; probe; diagnostic primer; component 1; ss.
XX
OS Banana bunchy top virus.
XX
PN US5756708-A.
XX
PD 26-MAY-1998.
XX
PF 24-FEB-1994; 94US-0202186.
XX
PR 24-FEB-1994; 94US-0202186.
XX
PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
XX
PI Burns TM, Dale JL, Harding RM, Karan M;
XX
DR WPI; 1998-321636/28.
XX
PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers and for producing virus-resistant plants
XX
PS Disclosure; Column 17-18; 59pp; English.
XX
CC This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plants or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
CC enhancers or termination signals.
XX
SQ Sequence 1111 BP; 363 A; 192 C; 276 G; 280 T; 0 other;

Query Match 14.1%; Score 59.6; DB 19; Length 1111;
Best Local Similarity 51.3%; Pred. No. 5.4e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 6 agtgcgtctgcgtatcgagagccagcgacgacgcacaaatcctctggatcgagcag 65
Db 584 agaggtggagaataatcattgcgcacccatgctcggagaataatttggctatggccc 643
QY 66 agacggagagagcggagtcggttcttggcaaatatctcggaactcaagcccgactggtt 125
Db 644 aaatggaggaaggaagaaacacgctatgcaaaacatctaataagacgagaaatgcgtt 703

```

Thu Oct 25 13:08:15 2001

QY 126 ctacacatgtgtggaaccagaaaggacgtattgtaccagtagacatcgagagcccaaaacg 185  
Db 704 ttattctccagagagaaatactatgtatgttagactgtataattacgaggatat--- 760  
QY 186 aaatttaactcctcgatgtaccagaggttaatttttagagattttaaatatgccctgttaga 245  
Db 761 ---tgttatatttgatatcccaagatgcaagaggattatttaaatcatgggttattaga 817  
QY 246 atgtgttaagaacagggcattcagttcggacaaataacgaccccttagttatcttgggtt 305  
Db 818 ggaatttaagaatggaataattcaaacgggaataatgaacccgttttgaagatag---t 874  
QY 306 cgaccatgtgcactgtctgtatttgccaatgtcctgcctgtattatttgaaaaatcagcag 365  
Db 875 agaatagtcgaagtcattgttaattgcttaacttcttcgcgaaggagaatctttctga 934  
QY 366 ggacagaataaaactg 381  
Db 935 agatogaataaagtgtg 950

Search completed: October 24, 2001, 10:07:12  
Job time: 617 sec



```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



Plate: 720 Row: D Column: 3  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 498.  
Location/Qualifiers  
1. .498  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-720 Col=3 Row=D"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites" 20 others

BASE COUNT  
171 a 107 c 87 g 113 t

Query Match 8.7%; Score 36.8; DB 231; Length 498;  
Best Local Similarity 56.7%; Pred. No.1.5;  
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 250 gttgaacaggcattcgctgcagcaaaatgcaacccttagtattctgtgggttcgac 309  
||| ||||||| ||||| ||||||| ||||| ||||| ||||| ||||| |||||  
Db 271 GTTCTGCCAGGCGCAATCAGGCNAGCACAAAGAATAAAGGGGTATTAAAGTTAGGAAAAGAG 330  
||| ||||||| ||||| ||||||| ||||| ||||| ||||| ||||| |||||  
QY 310 catgtgatgtactgcatttgccaatgtccctgcctgattttgaaaatacgcagggac 369  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 331 GAAGTCAAATTGCCCCTGTTTGCCAGATGACATGATTGTTTATTAGAAATCCCCCATCGAC 390  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3  
AW818171 518 bp mRNA EST 17-MAY-2000  
LOCUS CM1-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.  
DEFINITION AW818171  
ACCESSION AW818171 GI:7911165  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
Shogun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4&L2-CM1-ST0277-161)  
Seq primer: puc 18 forward  
High quality sequence start: 76  
High quality sequence stop: 183.  
Location/Qualifiers  
1. .518  
/organism="Homo sapiens"

| E-COLI DH10B <sup>a</sup> |       |      |          |
|---------------------------|-------|------|----------|
| SE COUNT                  | 77 a  | 66 c | 66 g     |
|                           | 77 a  | 66 c | 66 g     |
|                           | 108 t |      |          |
|                           |       |      | 1 others |

338 to

726 DGGCGCAACAGYKTADTGATATGKAKRSGAARGGWKTGRTCTRTARTATATGGDRKAGW 785

| BASE COUNT | 260 a | 123 c | 159 g | 217 t | 342 others |
|------------|-------|-------|-------|-------|------------|
| ORIGIN     |       |       |       |       |            |

726 DGGCGAAGAGYKTADTGATATGRKAKSGAARGWKGTGCTRTTATATATWGGDRKAGW 785

Thu Oct 25 13:08:18 2001

```

BASE COUNT      154 a      60 c      59 g      187 t
ORIGIN

Query Match      8.4%; Score 35.4; DB 104; Length 460;
Best Local Similarity 50.9%; Pred. No. 4;
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 223 tatttaattatgcctgttagaattgttaagaacagggcattcagttcggacaataac 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 TTTTGTGTTTACAAATTTGAAGTGTGTTTAAACAAAACAAATTAATTGTGCCAAAAA 64

QY 283 gaacccttagttatctgttggttcgaccatgtgcattgtactgtatttgccaattgctctg 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 AACCTCCATAATTATGTTGTTTAAAGCTCTATTATTAATCACTAGTATTGAAACAAACAGCAAG 124

QY 343 cctgattattgaaatcagcagggagacagaataaaactgtggaat 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 TATAATTTTGAATTCATCTAGTGAATTAATTAATTCATGTCGCAT 169

RESULT 8
BE603735 955 bp mRNA EST 02-MAR-2001
LOCUS HVSMEH0087003f Hordeum vulgare 5-45 DAP spike EST library
DEFINITION HVCDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0087003f,
mRNA sequence.
ACCESSION BE603735
VERSION BE603735.2 GI:13187835
KEYWORDS EST.
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare
REFERENCE 1 (bases 1 to 955)
AUTHORS Wing, R., Close, T. J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D. W., Main, D. and
Wood, T.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics
JOURNAL Unpublished (2000)
COMMENT On Aug 21, 2000 this sequence version replaced gi:9861296.
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCTCTACTAAAGG
High quality sequence stop: 931.
Location/Qualifiers
1. .955
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEH0087003f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
HVCDNA0009 (5 to 45 DAP)
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/notes="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      303 a      205 c      213 g      220 t
ORIGIN

Query Match      8.4%; Score 35.4; DB 137; Length 955;
Best Local Similarity 52.0%; Pred. No. 4.7;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 acggagagacgggaagtcggtgttttgccaaatatctcgactcaagcccgactgttct 127

```



```

Db 585 ACAGGGGACTGGGAACCTGGGTTACCAAACTAATCGCTTCAACACATCCCTTTTC 644
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 128 acacatgtgttggaacgaagaagacgtattgtaccagtcacatcgaggagcccaaacgaa 187
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 ACAGTTGGCGTATACAAAAGGCGAGCATTTCCCTTCCAAAAGTGCCCAATCTGAA 704
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 188 attaatcctcgtatgccagggtgtaatt 217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 705 GGGGAATGGGAATGGAGACGCTAATATT 734
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AZ401897/c
LOCUS
DEFINITION
  1M0168N22R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
  clone UUGCLM0168N22 R, DNA sequence.
VERSION
  AZ401897
KEYWORDS
  GSS.
SOURCE
  house mouse.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 345)
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  ,M., Rose,M., Rife,D., Stokes,R., Tingey,A., von Niederhausen,A.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0168 row: N column: 22
  Seq primer: CACACAGGAACACGTATGACC
  Class: plasmid ends
  High quality sequence stop: 345.
  Location/Qualifiers
    1..345
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGCLM0168N22"
      /clone_lib="Mouse 10kb plasmid UUGCLM library"
      /sex="Male"
      /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
      /notes="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnares/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA
      was blunt end-repaired with T4 DNA polymerase and T4
      polynucleotide kinase. Adaptor oligonucleotides were
      ligated to the blunt ends in high molar excess. The
      adapted DNA was purified and size-selected for a 9.5 to
      10.5 kb range using preparative agarose gel
      electrophoresis. Vector DNA was prepared from a derivative
      of PWD42 (g14732114/gb1AF129072.1), a copy-number
      inducible derivative of plasmid R1. The vector was ligated
      with adaptors complementary to the insert adaptors and
      purified. The sheared, adapted mouse DNA was annealed to
      adapted vector DNA, and transformed into
      chemically-competent E. coli XL10-Gold (Stratagene) cells
      and selected for ampicillin resistance."
FEATURES
  source
    1..345
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGCLM0168N22"
      /clone_lib="Mouse 10kb plasmid UUGCLM library"
      /sex="Male"
      /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
      /notes="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnares/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA
      was blunt end-repaired with T4 DNA polymerase and T4
      polynucleotide kinase. Adaptor oligonucleotides were
      ligated to the blunt ends in high molar excess. The
      adapted DNA was purified and size-selected for a 9.5 to
      10.5 kb range using preparative agarose gel
      electrophoresis. Vector DNA was prepared from a derivative
      of PWD42 (g14732114/gb1AF129072.1), a copy-number
      inducible derivative of plasmid R1. The vector was ligated
      with adaptors complementary to the insert adaptors and
      purified. The sheared, adapted mouse DNA was annealed to
      adapted vector DNA, and transformed into
      chemically-competent E. coli XL10-Gold (Stratagene) cells
      and selected for ampicillin resistance."

```

```

BASE COUNT      104 a      77 c      58 g      106 t
ORIGIN

Query Match
Best Local Similarity 8.3%; Score 35.2; DB 243; Length 345;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 216 tttagagtatttaattatgccctgttagaattgtttaagaacaggcatttcagttcgaa 275
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 TTTCAGAGCTTAAATGGGACGCCCTCTGAATGAGGTATAATATGCTATTACGTCATGA 271
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 276 caaatacgaaccccttagttatcttgggttcgaccatgtgcattgactctatttgcgaa 335
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 AACCTGAGGACATCTTTGTGGTGTAGTTGACTGCTGACTTGCTTGCCTTGCCTGTTGTC 211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
AQ746046
LOCUS
DEFINITION
  HS_2277_A2_F11.J7C CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=2277 Col=22 Row=K, DNA sequence.
ACCESSION
  AQ746046
VERSION
  AQ746046.1 GI:55233568
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 848)
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  99380589
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Clones may be purchased from Research Genetics (info@resgen.com).
  BAC end Web Server: http://www.htsc.washington.edu
  Plate: 2277 row: K column: 22
  Seq primer: T7
  Class: BAC ends
  High quality sequence stop: 848.
  Location/Qualifiers
    1..848
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="Plate=2277 Col=22 Row=K"
      /clone_lib="CIT Approved Human Genomic Sperm Library D"
      /sex="male"
      /notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
      E-Coli DH10B"
BASE COUNT      228 a      136 c      117 g      144 t      223 others
ORIGIN

Query Match
Best Local Similarity 8.3%; Score 35.2; DB 233; Length 848;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 250 gttagaacaggcatttcgagacaaatcgaaccccttagttatctctgggttcgac 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 GTTCTGCCAGGCAATCAGTAGGAGAAAGAAATAACCGGTATTCAATTAGGAAAGAG 393
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 310 catgtcactgtactgttgcgaatgtcctgcctgatttggaaatcagcaggagac 369
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





Thu Oct 25 13:08:18 2001

Query Match 8.0%; Score 33.8; DB 219; Length 604;  
Best Local Similarity 62.4%; Pred. No. 13;  
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
QY 178 ccaaaacgaaatttaactcgtatgtacccagggtgtaatttagagttatttaaaattatgcc 237  
Db 311 CTAAACAATTGAAGTGCACCTCTTAACACAGGTGAAATTTATAGTGTGCAAAATTATATA 370  
QY 238 ctgttagaatgtgttaagaacaggg 262  
Db 371 CTAATAAAACTGTTTAAAAAAAGG 395

Search completed: October 24, 2001, 13:13:40  
Job time: 11805 sec



|   |        |                    |           |              |
|---|--------|--------------------|-----------|--------------|
| Query Match   | 27.1%; | Score 114.2;       | DB 2;     | Length 1096; |
| Best Local Similarity   | 57.5%; | Pred. No. 2.5e-29; |           |              |
| Mismatches  | 0;     | Mismatches 163;    | Indels 3; | Gaps         |
| Matches   | 225;   | Conservative       |           |              |
| <p>19 atcgagagccagcgagcatcgcaacaatcctctgatatcgagcagagcggagagac 78</p> <p>551 ATCGAAGGTGTTCTTGATGATCGAAGATATCATCTGGGTATACGCTCCCAAGGAGCGGAA 610</p> |        |                    |           |              |

79 gggaagtcggtgttgccaaatattctgcgaaccacggcggcgccgcccacgagg-33  
||| ||||| | | | | | | | | | | | | | | | |  
611 GGAAAGTCAACCTTCGGAAGATATCTATCATTAACCGGATGGGATATATCAACGGT 670

671 GGAAGACGTCGGATATGATGCACATCATAACGATGGATCCGTGATAATCATTTGGATTATT 730

199 gatgtaccagggtglaatttagagtagtattaaattatgcctgtftagaagtgtttaagaac 258

731 CATATCCCCAGAGCAITTCAGATTCTCGAATTATGCGGTATTATAGAAACAATTAAGAAT 790

259 agggcatttcagttcggaacaatacaaac--ccctagttactcttggtctgcagccatgtg 315

791 AGAGTTTTAATAATACAAAATACGAACCATGCTGATTAGAAAAAGATGGACAAAATGTC 850

316 catgtactcgtatttgcgaatgtcctgcctgattatttgaataatcacgaggagcagagata 375

851 CATGTAAATTGTTATGGCAATGTGTGCCGTGATTATTGTAAAAATTCAGAAAGATAGAATA 910

376 aaactgtggaatatattaagtagtgtgtcatc 406

911 AAATAATTAATCTTCGAGAAAGAAACTTC 941

RESULT 2  
US-08-418-071-5  
Sequence 5, Application US/08418071  
Patent No. 5846705  
GENERAL INFORMATION:  
APPLICANT: Wu, Rey-Yuh  
APPLICANT: You, Li-Ru

TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTION OF BANANA BUNCHY TOP VIRUS  
 TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue Of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10020

COMPUTER READABLE FORM. disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: word Perfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/418,071  
FILING DATE: 06-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: DCB-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1091 base pairs  
TYPE: nucleic acid



| Query Match           | 20.9%        | Score 88.4   | DB 4           | Length 1022 |        |
|-----------------------|--------------|--|----------------|-------------|--------|
| Best Local Similarity | 54.3%        | Pred. NO. 1.4e-20  |                |             |        |
| Matches 201           | Conservative | 0  | Mismatches 166 | Indels 3    | Gaps 1 |
| yy                    | 25           | gagccagcgagcattcgcaaatcctctgatatcgcgacgagcagcgagcagcgaggaag    | 84             |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 559          | gcggaaaccagatcgagctatcatctcggtctatggttcacagcgaggaagaaag        | 618            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 85           | tcogtgttgccaaatatctcggaactcaagccogactggtttctacaatgagtgaacc     | 144            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 619          | acgagcttcgcgaaggaatt--aatcagctatgagatggttttatcacgagggaggaag    | 675            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 145          | agaaagacgtattgtaccagctacatcgagcaccacaaacgaatttaactcctcgatga    | 204            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 676          | accacagacggtattatattatgtatgtgtcagacccacagaggaataatccggttgatggt | 735            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 205          | cccaggtgtaatttagagatttaaaattgcctctttagaatggttgaagacacggcca     | 264            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 736          | cccagggtgtcttcggagatgatgaactatcagcgcattggagatgttgaaagacagagtt  | 795            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 265          | ttcagttcggacaaatcgaaccccttagtactctcgtgggttcgacctgctgaatgtactc  | 324            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 796          | tttgcaagtacaaaattatagcccttgtagatctttgtattaggaagttagttcatttaatt | 855            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 325          | gtattggcaatgtcctgcctgattatttgaaatcagcagggacagaaataaaactcgtgg   | 384            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 856          | gtggttgccaacgtggcacccttgaccgccgcaataagtgaggacacagacttgtaattatc | 915            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 385          | aatatttaa  | 394            |             |        |
| bb                    |              |  |                |             |        |
| yy                    | 916          | aatgttgtaa   | 925            |             |        |
| bb                    |              |  |                |             |        |

```

; Sequence 14, Application US/08202186
; Patent NO. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Toley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202.186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
;
; JS-08-202-186-14

```

|                       |   |              |               |            |      |        |       |
|-----------------------|---|--------------|---------------|------------|------|--------|-------|
| Query Match           | 14.9%;  | Score        | 62.8;         | DB         | 1;   | Length | 1110; |
| Best Local Similarity | 53.2%;  | Pred.        | No. 7, 7e-12; |            |      |        |       |
| Matches               | 183;  | Conservative | 0;            | Mismatches | 155; | Indels | 9;    |
| Gaps                  |   |              |               |            |      |        |       |
| 38                    | atcgcacaactctctgatatgcgacgacgacgaggaacagcggaagtcctgtttgccg    | 97           |               |            |      |        |       |
| 615                   | ATCGAAGAATTTATTTGGGTTTACGGGCCAATATCGAGGAGGAAGAAACGACGTATGCAA  | 674          |               |            |      |        |       |
| 98                    | aatctcggactcaagccgactggtcttcacacatggtgtggaaccagaaaggacgtat    | 157          |               |            |      |        |       |
| 675                   | ACATCTCAATCGACACGAGNATCGGTTTATTCTCCAGGAGGAAATCATTTGGATATAT    | 734          |               |            |      |        |       |
| 158                   | tgtaccaggtacatcgaggaccocaaaaacgaataatctctcgatgtaccaggtgaatt   | 217          |               |            |      |        |       |
| 735                   | GTAGACTGTATAATTATGAGGATAT-----TGTATATTATTGATATCCCTACATGCAAG   | 788          |               |            |      |        |       |
| 218                   | tagagtatttaaatatgccctgttagaagtgtttaagaacagggcattcagttcggaca   | 277          |               |            |      |        |       |
| 789                   | AGATTATTTAAATATCGTTTATTAGAGGAATTTAAGAAATGCAATAATCAAGCGGA      | 848          |               |            |      |        |       |
| 278                   | aatacgaaccccttagttatcttgggttcgaccatgtcagtcgatttgcgcaatg       | 337          |               |            |      |        |       |
| 849                   | AATATGAACCCGTTTGTGAAGATG---TAGAATATGTGGAAGTCATTGTAAATGGCTTACT | 905          |               |            |      |        |       |
| 338                   | tctgcctgattatttgaataatcagcagggcacagataaaaactg                 | 381          |               |            |      |        |       |

RESULT 7  
US-08-973-068-28  
: sequence 28. Application US/08973068









APPLICANT: DALE, James L.  
 APPLICANT: HARDING, Robert M.  
 TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
 NUMBER OF INVENTIONS: 60  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W.  
 CITY: Washington, D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/202,186  
 FILING DATE: 24-FEB-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: JEFFERY, Donald D.  
 REGISTRATION NUMBER: 19,980  
 REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202 672 5300  
 TELEFAX: 202 672 5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1105 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single

```

Query Match          13.9%; Score 58.6; DB 1; Length 1105;
Best Local Similarity 52.5%; Pred No. 2.1e-10;
Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;

Qy 38 atcgacacaatcctctggatagtcgagacagagagagagagagcgtggtttgcc 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 ACCGAGGATAATTTGGGTCTATGGCCCAATGGAGGAGGAAGAACACTTATGC 677

Qy 98 aatctctcggaactcaagccgactggtctcacacatgctgtggaacagaaaggcgtat 157
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 678 AATATTTAATGAACACGAAGATGCGTTTTATTTCGCCAGGAGGAAATCATTTGG 737

Qy 158 tgcacgagtcacatcgaggaccacaaacgaataattacctcgatgtaccagggtgaatt 217
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 738 GTAGATTGTATAATTAATGAGGAATA-----GTTATATTGTATATCCAGATCG 791

Qy 218 tagagtatattaataatagccctgttaagatgtttaagaacagggcattcagtcgcga 277
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 792 AGGAATATTTAAACTATCGGTTTATTAGACAAATTTAGAAATGAAATTAITCAAGCG 851

Qy 278 aatacgaaccccttagttatcttgggtctgcacctgctgcactgctatctgtattgc 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 852 AATATGAACCCGTTTGTGAAATTTG-----TAGAATATGTGGAAGTCATTGTA 908

Qy 338 tctgcctgattatttgaataatcagcagggcagagataaaact 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 909 TCCTTCCGAAGGAAGGAATCTTTTCAGAGATCGAATAAAGCT 951

RESULT 14
US-08-202-196-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.

```

| Query Match           | 13.74;       | Score 58;   | DB 1;           | Length 1111; |
|-----------------------|--------------|---|-----------------|--------------|
| Best Local Similarity | 51.1%;       | Pred. No. 3.3e-10;  |                 |              |
| Matches 192;          | Conservative | 0;  | Mismatches 175; | Indels       |
|                       |              |   |                 | Gaps         |
| QY                    | 6            | agtgctgtctcgatcggagagccgagcgagcgtcgcacaaatcctctgatatcgcgagc     | 65              |              |
|                       |              |   |                 |              |
| Db                    | 584          | AGAGGTGGAGATAATCATCGCGCCAGCCATGTCATCGGAGAAATAATTGGTGCTATGGGCC   | 643             |              |
| QY                    | 66           | agacggagggagcgggaagtcggtgtttgcgaatatctcggactcaagcccgactggtt     | 125             |              |
|                       |              |   |                 |              |
| Db                    | 644          | AAATGGAGGAGGAAGAAAGACAACGTTATGCAAAACATCTAATCAACAGCAGGAATCGCGTT  | 703             |              |
| QY                    | 126          | ctacacatgtgtggaaccgaaaggacgtatgtaccggtacatcgaggaccacaaacg       | 185             |              |
|                       |              |   |                 |              |
| Db                    | 704          | TTATTCTCCAGGAGGAANAATCTGGGATATATGCTAGACTGTATAATTACGAGGATAT---   | 760             |              |
| QY                    | 186          | aaatttaactctcgtatgtaccaggggtgaattttagsgatttaaattatgccctgttaga   | 245             |              |
|                       |              |   |                 |              |
| Db                    | 761          | ---TCTTATATTGATATTCCAAGATGCAAGAGGATATTTTAAATTAATGGGTATTATGA     | 817             |              |
| QY                    | 246          | atggtttaagaaacaggcctcagtcgttcgacaaatacgaaccccttagttatcttggggt   | 305             |              |
|                       |              |   |                 |              |
| Db                    | 818          | GGAAATTTAAGAATGGAATAATTCAAACGCGGAATATGAACCCGTTTGAAGATAG--T      | 874             |              |
| QY                    | 306          | cgaccatctgcatgctactcgtattttgccaatgctcgtcgtgattatttgaaaaatcagcag | 365             |              |
|                       |              |   |                 |              |
| Db                    | 875          | AGAATATGTCGAAGTCATCTGTAATGCGTAACCTTCCTCCGAAGGAAGAAATCTTTCCTGA   | 934             |              |
| QY                    | 366          | ggacagataaaaactg  | 381             |              |
|                       |              |   |                 |              |
| Db                    | 935          | AGATCGAATATAGTTG  | 950             |              |

Search completed: October 24, 2001, 10:00:16

us-09-462-955-1\_copy\_583\_1004.rni

---

Thu Oct 25 13:08:17 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:12 ; Search time 393.52 Seconds  
(without alignments)  
1329.137 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_383\_1215  
Perfect score: 833  
Sequence: 1 ggggtctctcagcagcagga.....accatccaggtagtagct 833

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1450202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 120.6 | 14.5        | 1096   | 20 | AAV71833    |
| 2          | 120.6 | 14.5        | 1096   | 21 | AAA38950    |
| 3          | 117   | 14.0        | 1091   | 20 | AAV71834    |
| 4          | 117   | 14.0        | 1091   | 21 | AAA38947    |
| 5          | 115.6 | 13.9        | 886    | 20 | AAV71832    |
| 6          | 102.2 | 12.3        | 1106   | 20 | AAV71832    |
| 7          | 100.6 | 12.1        | 1106   | 21 | AAA38946    |
| 8          | 99.2  | 11.9        | 1017   | 17 | AAV71832    |
| 9          | 91.2  | 10.9        | 1022   | 17 | AAV71832    |
| 10         | 81.2  | 9.7         | 593    | 20 | AAA34686    |
| 11         | 62.8  | 7.5         | 1110   | 19 | AAV24089    |

|    |      |     |       |    |          |                    |
|----|------|-----|-------|----|----------|--------------------|
| 12 | 59.6 | 7.2 | 982   | 18 | AAV49405 | Banana bunchy top  |
| 13 | 59.6 | 7.2 | 1110  | 19 | AAV24086 | Banana bunchy top  |
| 14 | 59.6 | 7.2 | 1111  | 19 | AAV24077 | Banana bunchy top  |
| 15 | 59.6 | 7.2 | 1111  | 19 | AAV24084 | Banana bunchy top  |
| 16 | 59.6 | 7.2 | 1111  | 19 | AAV24087 | Banana bunchy top  |
| 17 | 58.6 | 7.0 | 1103  | 19 | AAV24091 | Banana bunchy top  |
| 18 | 58.6 | 7.0 | 1105  | 19 | AAV24093 | Banana bunchy top  |
| 19 | 58.6 | 7.0 | 1109  | 19 | AAV24088 | Banana bunchy top  |
| 20 | 58   | 7.0 | 1111  | 19 | AAV24090 | Banana bunchy top  |
| 21 | 57   | 6.8 | 1104  | 19 | AAV24092 | Banana bunchy top  |
| 22 | 56.4 | 6.8 | 1111  | 19 | AAV24085 | Banana bunchy top  |
| 23 | 49.4 | 5.9 | 936   | 22 | AAF58252 | Oligonucleotide D1 |
| 24 | 49.4 | 5.9 | 936   | 22 | AAF58252 | Oligonucleotide D1 |
| 25 | 49.4 | 5.9 | 936   | 22 | AAF58254 | Oligonucleotide D1 |
| 26 | 49.4 | 5.9 | 936   | 22 | AAF58254 | Oligonucleotide D1 |
| 27 | 49.4 | 5.9 | 936   | 22 | AAF58257 | Oligonucleotide D1 |
| 28 | 49.4 | 5.9 | 936   | 22 | AAF58257 | Oligonucleotide D1 |
| 29 | 49.4 | 5.9 | 936   | 22 | AAF58259 | Oligonucleotide D1 |
| 30 | 49.4 | 5.9 | 936   | 22 | AAF58262 | Oligonucleotide D2 |
| 31 | 49.4 | 5.9 | 936   | 22 | AAF58262 | Oligonucleotide D2 |
| 32 | 49.4 | 5.9 | 938   | 22 | AAF58255 | Oligonucleotide D2 |
| 33 | 49.4 | 5.9 | 938   | 22 | AAF58255 | Oligonucleotide D1 |
| 34 | 49.4 | 5.9 | 938   | 22 | AAF58255 | Oligonucleotide D1 |
| 35 | 46.4 | 5.6 | 287   | 20 | AAV71830 | Subgenomic fragmen |
| 36 | 46.4 | 5.6 | 287   | 21 | AAA38948 | Banana bunchy top  |
| 37 | 46.4 | 5.6 | 300   | 20 | AAV71831 | Subgenomic fragmen |
| 38 | 46.4 | 5.6 | 300   | 21 | AAA38949 | Banana bunchy top  |
| 39 | 38   | 4.6 | 244   | 22 | AAF58238 | Oligonucleotide D1 |
| 40 | 38   | 4.6 | 244   | 22 | AAF58238 | Oligonucleotide D1 |
| 41 | 33   | 4.0 | 9345  | 21 | AAZ93966 | Mouse uromodulin p |
| 42 | 32.6 | 3.9 | 5053  | 19 | AAV22816 | Nucleotide sequenc |
| 43 | 32.6 | 3.9 | 5055  | 19 | AAV17236 | DNA from a region  |
| 44 | 31.4 | 3.8 | 41599 | 21 | AAZ35351 | Cosmid including s |
| 45 | 31.2 | 3.7 | 5692  | 22 | AAF32248 | Streptomyces sp. c |

## ALIGNMENTS

RESULT 1  
ID AAV71833 standard; DNA; 1096 Bp.  
XX AC AAV71833;  
XX AC AAV71833;  
XX 10-FEB-1999 (first entry)  
XX BBTV DNA II clone (2-17) nucleotide sequence.  
DE BBTV DNA II clone (2-17) nucleotide sequence.  
XX Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
XX Banana bunchy top virus.  
XX Banana bunchy top virus.  
FH Key Location/Qualifiers  
FT TATA\_signal 1..7  
FT stem\_loop /\*tag= a  
FT /\*tag= 8..38  
FT /\*tag= b  
FT CDS 70..928  
FT /\*tag= c  
FT /\*product= "ORF-V2 product"  
FT /\*transl\_except= (pos:215..217, aa:Gly)  
FT polyA\_signal 533..538  
FT /\*tag= d  
FT polyA\_signal 799..804  
FT /\*tag= e  
FT polyA\_signal 907..912  
FT /\*tag= f  
FT polyA\_signal 1030..1035  
FT /\*tag= ge  
XX US5846705-A.







30-JUN-1994; 94TW-0106105.  
30-JUN-1994; 94TW-0106105.  
(BIOT-) DEV CENT BIOTECHNOLOGY.  
Wu R, You L, Song T;  
WPI; 2000-316145/27.  
Two circular single-stranded DNAs associated with banana bunchy top  
virus and detection of the virus -  
Claim 2; Page 1; 7pp; Chinese.  
The present invention describes two circular single-stranded DNAs  
associated with banana bunchy top virus (BBRV) and the encoded proteins  
Also described is a method of detecting the BBRV virus using the  
polymerase chain reaction (PCR). The present sequence represents a  
BBRV related nucleotide sequence from the present invention.  
sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;

|                       |              |                    |                 |                    |
|-----------------------|--------------|--------------------|-----------------|--------------------|
| Query Match           | 14.0%;       | Score 117;         | DB 21;          | Length 1091;       |
| Best Local Similarity | 52.2%;       | Pred. No. 2.8e-29; |                 |                    |
| Mismatches 314:       | Conservative | 0;                 | Mismatches 275; | Indels 12; Caps 2; |

[illegible]

|    |   |  |
|----|---|--|
| XX | RESULT 5  |  |
| XX | AAAX34687   |  |
| XX | ID AAX34687 standard; DNA; 886 BP.  |  |
| XX | AAAX34687;  |  |
| XX | 02-JUL-1999 (first entry)   |  |
| XX | DNA sequence of BBTv S2I and S2L promoter fragments.                      |  |
| XX | Promoter: Banana Bunchy Top Virus; BBTv; gene transcription; ubil1;       |  |
| XX | polyubiquitin 1; banana; SS.  |  |
| XX | Banana bunchy top virus.  |  |
| XX | WO9915646-A1.   |  |
| XX | 01-APR-1999.  |  |
| XX | 21-SEP-1998; 98WO-AU00786.  |  |
| XX | 30-JUN-1998; 98AU-0004423.  |  |
| XX | 19-SEP-1997; 97AU-0009339.  |  |
| XX | (UYQU-) UNIV QUEENSLAND TECHNOLOGY.                                       |  |
| XX | Becker DK, Dale JL, Dugdale B, Harding RM, Hermann SR;                    |  |
| XX | WPI; 1999-254706/21.  |  |
| XX | Banana Bunchy Top Virus promoter constructs                               |  |
| XX | Examples; Fig 23; 84pp; English.  |  |
| XX | The invention relates to DNA promoter sequences derived from components   |  |
| XX | of Banana Bunchy Top Virus (BBTV). The DNA molecule includes a promoter   |  |
| XX | sequence derived from an untranslated portion of any one BBTv component   |  |
| XX | and is adaptable for promoting transcription of a cloned gene in a plant  |  |
| XX | cell. The promoter, included in a DNA chimeric vector, is useful for the  |  |
| XX | expression of a gene in a plant cell. The inclusion of the polyubiquitin  |  |
| XX | 1 (ubil1) intron sequence into BBTv promoter constructs enhances promoter |  |
| XX | activity. The present sequence represents the DNA sequence of BBTv S2I    |  |
| XX | promoter and S2L promoter fragments.                                      |  |
| XX | Sequence 886 BP; 294 A; 171 C; 199 G; 222 T; 0 other;                     |  |
| XX | SO  |  |

|                       |              |                    |                 |             |
|-----------------------|--------------|--------------------|-----------------|-------------|
| Query Match           | 13.9%;       | Score 115.6;       | DB 20;          | Length 886; |
| Best Local Similarity | 52.4%;       | Pred. No. 7.5e-29; |                 |             |
| Matches 309;          | Conservative | 0;                 | Mismatches 269; | Indels 12;  |
|                       |              |                    |                 | Gaps 2;     |

[illegible]

Db 360 cattaaaacctggtggtggtatatacaacggtggaagagctggatgatgacatca 419  
 QY 367 acatcgagaccacaaacgaaattatcccgatgtaccaggtgtaatttagagtatt 426  
 Db 420 taacgatggtcgtgataatcatggtattgatatcccaagtcattcgattatc 479  
 QY 427 taattatgcccgtgtagaattgtttaagaacagcggtcattcagttcggacaaatacgaac 486  
 Db 480 tgaattatggtttatagaacaaataaagaatagagttttaataatacaaaatacgaac 539  
 QY 487 c---ccttagttatcttggttgcacacctgtgcatgtactctgtatttgcaatgtctgc 543  
 Db 540 catgtgtgattagaaagatggacaaatgtccatgttaattgttatggcaaatgtgttc 599  
 QY 544 ctgattattgaaatcagcaggacagacagaataaaactgtggaaatttta 593  
 Db 600 ctgattattgaaatttcagagatagaataaaaaataatttaattgtga 649

RESULT 6  
 AAV71832  
 XX AAV71832 standard; DNA; 1106 BP.  
 AC  
 XX  
 XX  
 DT 10-FEB-1999 (first entry)  
 XX  
 XX  
 DE BBTV DNA I clone (7-4-2) nucleotide sequence.  
 XX  
 KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
 KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
 XX  
 OS Banana bunchy top virus.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT TATA\_signal 1..8  
 FT /\*tag= a  
 FT stem\_loop 9..38  
 FT /\*tag= b  
 FT CDS 62..922  
 FT /\*tag= c  
 FT /product= "ORF-V2 product"  
 FT /transl\_except= (pos:335..337, aa:Gly)  
 FT /transl\_except= (pos:518..520, aa:Ser)  
 FT polyA\_signal 380..385  
 FT /\*tag= d  
 FT polyA\_signal 901..906  
 FT /\*tag= e  
 XX  
 PN US5846705-A.  
 XX  
 XX 08-DEC-1998.  
 XX  
 PF 06-APR-1995; 95US-0418071.  
 XX  
 PR 06-APR-1995; 95US-0418071.  
 XX  
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.  
 XX  
 PI Soong T, Wu R, You L;  
 XX  
 DR WPI; 1999-059037/05.  
 DR P-PSDB; AAW87459.  
 XX  
 PT Nucleic acids having banana bunchy top virus component sequences -  
 PT used to design primers for use in polymerase chain reaction  
 PT detection of the virus  
 XX  
 XX  
 PS Claim 1; Fig 11A-B; 27pp; English.  
 XX  
 CC This represents the nucleotide sequence of a banana bunchy top virus  
 CC (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The  
 CC invention provides nucleic acid sequences associated with BBTV that can

CC be used in a PCR technique for detecting BBTV. The nucleic acid  
 CC sequences (AAV71830 to AAV71833) are used as the basis for the  
 CC construction of PCR primers, to detect BBTV infection. The PCR technique  
 CC is used for detecting BBTV in plant tissues (preferably banana,  
 CC especially Musa species). The virus, one of the most important banana  
 CC species viruses, causes phloem damage and is transmitted by aphids. PCR  
 CC detection gives accurate, reliable and specific determination of absence  
 CC or presence of the virus.  
 XX

SQ Sequence 1106 BP; 335 A; 225 C; 257 G; 289 T; 0 other;

Query Match 12.3%; Score 102.2; DB 20; Length 1106;  
 Best Local Similarity 53.3%; Pred. No. 3e-24;  
 Matches 215; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 195 tggcagcttgaagtgtctgctgcgagcgagccgagcgagcgatcgcaaatccctctg 254  
 Db 524 tggcagattcaattgacggaggccaattgacgaggaaccgagatcggaagcatctcgg 583  
 QY 255 atatcgcgagcagcggagcggagcggagcggagcggagcggagcggagcggagcggag 314  
 Db 584 gtctatggtctcttattggttaagtgaaggtaatacaacatcgcaagtcactaatcaagaag 643  
 QY 315 ccgagactggtctacacatgtggtggaaccgagaaagcgtattgtaccagtcacatcgag 374  
 Db 644 gattggtctctacaccggggtggaagaggaatacttattctctctatgtgagcga 703  
 QY 375 gacccaaaaacgaatttaactctcgtatgacccaggtgtaatttagagatttaaatatt 434  
 Db 704 ggaatcgacaaacataatgattgatactctctgttgaatcaggattatttaaatatt 763  
 QY 435 gccctttagaatgtgttaagaacagggcggttcagttcggacaaatcgacaccccttagt 494  
 Db 764 gatgtaatagaggcattaaaggatagggttatagagagactactaaatacaaacccataaag 823  
 QY 495 tatcttgggttcgaccatgtgcattactctgattgttggcgaatgctcgtcgtgattttg 554  
 Db 824 atagttgaattaggttaaatcatgctgtaactgcatggtggaatttcgctgactctgt 883  
 QY 555 aaatcagcagggacagaataaaactgtggaatttttaaaagta 597  
 Db 884 aaatctccgaagatcgataaaataaaatcatttattgctgaagaa 926

RESULT 7  
 AAA38946  
 ID AAA38946 standard; DNA; 1106 BP.  
 XX  
 AC AAA38946;  
 XX  
 XX  
 DT 23-AUG-2000 (first entry)  
 XX  
 DE Banana bunchy top virus related nucleotide sequence #1.  
 XX  
 KW Banana bunchy top virus; BBTV; detection; ds.  
 XX  
 OS Banana bunchy top virus.  
 XX  
 PN TW360710-A.  
 XX  
 PD 11-JUN-1999.  
 XX  
 PF 30-JUN-1994; 94TW-0106105.  
 XX  
 PR 30-JUN-1994; 94TW-0106105.  
 XX  
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.  
 XX  
 PI Wu R, You L, Song T;  
 XX  
 DR WPI; 2000-316145/27.  
 XX

Two circular single-stranded DNAs associated with banana bunchy top virus and detection of the virus -  
 Claim 1; Page 1; 7pp; Chinese.  
 The present invention describes two circular single-stranded DNAs associated with banana bunchy top virus (BBTV) and the encoded proteins. Also described is a method of detecting the BBTV virus using the polymerase chain reaction (PCR). The present sequence represents a BBTV related nucleotide sequence from the present invention.  
 sequence 1106 BP; 335 A; 226 C; 256 G; 289 T; 0 other;

|     | Query Match           | 12.1%   | Score 100.6;       | DB 21;          | Length 1106; |
|-----|-----------------------|---|--------------------|-----------------|--------------|
|     | Best Local Similarity | 53.1%;  | Pred. No. 1.1e-23; |                 |              |
|     | Matches 214;          | Conservative  | 0;                 | Mismatches 189; | Indels 0;    |
|     |                       |   |                    |                 | Gaps 0;      |
| QY  | 195                   | tggcagcttgaagtgcgtctgcgcacgcgagaccgcgagcagcgcacaaatcctctgg        | 254                |                 |              |
|     |                       |   |                    |                 |              |
| DBb | 524                   | tggcagattcaattgacgcggaggaacattgacgaaaccgcagtgatcgaaagcatcctgg     | 583                |                 |              |
|     |                       |   |                    |                 |              |
| QY  | 255                   | atatgcggacgacgacgagcagcgggaaatccgctgtttggcaaatatctcggactcaag      | 314                |                 |              |
|     |                       |   |                    |                 |              |
| DBb | 584                   | gctatggtccttatggttaatgaggtaaatcaacatatgcgaagtcaataatcaagaag       | 643                |                 |              |
|     |                       |   |                    |                 |              |
| QY  | 315                   | cccgactggctctacacatgtggttgggaaccagaaaaggacgctattgtaccagtcacatcgag | 374                |                 |              |
|     |                       |   |                    |                 |              |
| DBb | 644                   | gattggtctcacacaggggtgggaagaaggaaatatcttatctcctatgtggacgaa         | 703                |                 |              |
|     |                       |   |                    |                 |              |
| QY  | 375                   | gacccaaacgaaattaatcctcgatgtacccagggtgtaatttagagttattaaattat       | 434                |                 |              |
|     |                       |   |                    |                 |              |
| DBb | 704                   | ggaatgcacagcatatgatttgattctcctcttgtaatcaggattattttaaattat         | 763                |                 |              |
|     |                       |   |                    |                 |              |
| QY  | 435                   | ggcctgttagaatgtgttaagaacaggcgattcagtcggacaaaatacgaacccctagt       | 494                |                 |              |
|     |                       |   |                    |                 |              |
| DBb | 764                   | gatgtaataagagcattaaaggatagggttatagagagtgactaaatacaaacccataag      | 823                |                 |              |
|     |                       |   |                    |                 |              |
| QY  | 495                   | tattctgggttcgaccatgtgcagtactcgtattttccaaagtctcctgcctgattatttg     | 554                |                 |              |
|     |                       |   |                    |                 |              |
| DBb | 824                   | atagttggaattaggtaataacatgtaatcgtcagtcgaatttcagtcgctgactctgt       | 883                |                 |              |
|     |                       |   |                    |                 |              |
| QY  | 555                   | aaaatcagcaggcagacataaaactgtggaaattttaaagta                        | 597                |                 |              |
|     |                       |   |                    |                 |              |
| DBb | 884                   | aaaatcfcggaagatcgaataaaaatcattttatgctgaagaa                       | 926                |                 |              |
|     |                       |   |                    |                 |              |

|          |  |                     |
|----------|--|---------------------|
| RESULT   | 8  |                     |
| AAT13165 |  |                     |
| ID       | AAT13165 standard; DNA; 1017 BP.                         |                     |
| XX       |  |                     |
| XX       | AAT13165;  |                     |
| XX       |  |                     |
| XX       | 23-MAY-1996 (first entry)                                |                     |
| DT       |  |                     |
| XX       |  |                     |
| XX       | SCSV segment 6.  |                     |
| DE       |  |                     |
| XX       |  |                     |
| XX       | SCSV; promoter; transcription; transgenic plant; legume; |                     |
| KW       | gene expression; crop improvement; ss.                   |                     |
| KW       |  |                     |
| XX       |  |                     |
| XX       | Subterranean clover stunt virus isolate F.               |                     |
| OS       |  |                     |
| XX       |  |                     |
| XX       |  |                     |
| FH       | Key  | Location/Qualifiers |
| FT       | CDS  | 48..905             |
| FT       |  | /*tag= a            |
| XX       |  |                     |
| XX       |  |                     |
| XX       | WO9606932-A1.  |                     |
| XX       |  |                     |
| XX       | 07-MAR-1996.   |                     |
| PD       |  |                     |
| XX       |  |                     |
| XX       | 30-AUG-1995;   | 95WO-AU00552.       |
| PF       |  |                     |
| XX       |  |                     |
| XX       | 07-NOV-1994;   | 94AU-0009281.       |
| XX       |  |                     |
| PR       |  |                     |

|     |  |               |
|-----|--|---------------|
| PR  | 30-AUG-1994;   | 94AU-0007770. |
| XX  | (CSIR ) COMMONWEALTH SCI & IND RES ORG.                                |               |
| XX  | (AUSU ) UNIV AUSTRALIAN NAT.   |               |
| XX  | Boevink PC, Chu PMG, Keese PK, Khan RI, Larkin PJ;                     |               |
| PI  | Marshall JS, Surin BP, Taylor WC, Waterhouse PM;                       |               |
| PI  | WPI: 1996-160363/16.   |               |
| DR  | Circovirus transcription regulatory sequences and related constructs   |               |
| XX  | - useful in plants, esp. leguminous plants, for the modulation of      |               |
| PPT | gene expression  |               |
| PT  | Claim 4; Page 62; 121pp; English.                                      |               |
| XX  | The subterranean clover stunt virus (SCSV) genome has at least 7       |               |
| XX  | distinct ssDNA components, designated segments 1-7 (AAR13160-66), each |               |
| CC  | contg. 1 major open reading frame and a non-coding region. Segment     |               |
| CC  | 6 is predicted to be a viral replication-associated protein gene.      |               |
| CC  | Genetic constructs useful in the genetic engineering of plants (esp.   |               |
| CC  | legumes) comprise 1 or more heterologous gene(s) operatively linked    |               |
| CC  | to a promoter region, and in some cases also a terminator region,      |               |
| CC  | selected from segments 1-7. The transcription regulators facilitate    |               |
| CC  | expression of foreign genes in plants and also facilitate control of   |               |
| CC  | levels of gene expression in different plant tissue types.             |               |
| XX  | Sequence 1017 BP: 312 A; 160 C; 265 G; 280 T; 0 other;                 |               |
| XX  |  |               |

```

Query Match      11.98; Score 99.2; DB 17; Length 1017;
Best Local Similarity 54.1%; Pred. No. 3e-23;
Matches 225; Conservative 0; Mismatches 188; Indels 3; Gaps 1;

Qy 189 cacaaattgcagcttgaagtctgtctgcgcatcggagagccagcggaocgatcgcacaatc 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 cgcccttgcaagaagtgatgcttctgttagaggagaagaccagattatagaacgata 557

Qy 249 ctctgatatcggcagcagagacggaggagacgggaagtcctgtttgccaaatatctcgga 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 atctgggtgtatggaccctcgtggttaatgaaaggccaaatctacatttgcagaacatctgtca 617

Qy 309 ctcaagcccgactggtctcacatgtggttggaaccagaaagagacgtattgtaccagttac 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 ttgaagaatggttgggtttatctgcgtgaggaaagacacaagatatgatgcattctgtg 677

Qy 369 atcgaggaccacaaacgaaatttaactctcgatgtaccacgggtgtaatttagagtattta 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 actgcgcgccttaagaataattgggtatttgacatacccagagttagttccagagtattgtg 737

Qy 429 aattatgcccctgttagaatgttgaagaacaggcgattcagttcggacaaatcacgaacc- 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 aattatggtgtaatagaacacaggtttaagaatagggttaatgtgtaactaactagttagacca 797

Qy 488 --ccctagttatcttgggttcgaccatgtgcatgtactcgtatttggccaatgctcctgcct 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 tgtgtaatgcggatgataatactcatcctgttccatgtaattgtgttggaagtactcccca 857

Qy 546 gattatttgaaatcagcgggacagagataaaactgctggaatatattaaagtattgtg 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 858 gatttgggaaaattaaagtgaaagatagaataaattaaattcgtttgttgaaacictg 913

RESULT 9
AAT13161
ID AAT13161 standard; DNA; 1022 BP.
XX
AC AAT13161;
XX
XX 23-MAY-1996 (first entry)
XX
XX SCSS segment 2.
XX
XX

```





|    |     |   |     |
|----|-----|---|-----|
| Db | 704 | ttatttccagaggaaaaatcattggatactagactgtataattacaggatat---         | 760 |
| QY | 386 | aaatttaattcctcgatgtaccagggtgtaatttagagttatttaaatattacctgttaga   | 445 |
| Db | 761 | ---tggtattatttgatctccaagatgcaagaggattatttaattatgggttattaga      | 817 |
| QY | 446 | atggtgttaagaacagggcattccgcttggaacaaatacgaaccccttagttatctgggtt   | 505 |
| Db | 818 | ggaatttaagaaatggaaataatccaagcgggaataatgaacccgttttgaagatag---    | 874 |
| QY | 506 | cgacctgtgcattgtactcgttatttgcgaatgtcctgcctgattattgaaatacagcag    | 565 |
| Db | 875 | agaataatgcgaagtcattgttaatggctaaactttctccgaaggagaagaaatttttcctga | 934 |
| QY | 566 | ggacagaataaaaactg   | 581 |
| Db | 935 | agatcgaataaaagt   | 950 |

RESULT 14

Example 1; Fig 11; 59pp; English.

Query Match 7.2%; Score 59.6; DB 19;  
Best Local Similarity 51.3%; Pred. NO. 8.8e-10;  
Matches 193; Conservative 0; Mismatches 174; Indexes q.

Qy 206 agtgcgtctgcgatcggagagccagcgacgatcgacaatctcttgatatcgcgagc 265

Qy 206 agtgcgtctgcgcatcgagagagccagcgacacatcgacacatccctctggtatgcggacg 265  
 Db 584 agaggtggagaaatcatcgccagccatgctcatcgagagaaataatttgggtctatggccc 643  
 Qy 266 agacggagagagcggaagtcccggtttgccaataatctcggactcaagcccgactgggtt 325  
 Db 671 aaatggagagagagaaacacgctatgcaaaacatctaatgaagacgagaaatgcgtt 703  
 Qy 326 ctacacatgtgtggagaccagaaagacgtattgtaccagtagcatcgagaccccaaacg 385  
 Db 704 ttattctccaggagaaatcattggatatatgtagactgtataattacaggatatat--- 760  
 Qy 386 aaatttaactcgtatgcacgggtgtaatttagagatttttaaaattatgcctgttaga 445  
 Db 761 ---tggtattatgtatattccaagatgcaaaagaggtatttttaaaattatgggttattaga 817  
 Qy 446 atgtgttaagaacacagggcattcagttcggaacaaatacgaaccccttagttatcttgggtt 505  
 Db 818 ggaatttaagaatggaataattcaaacggggaataatgaacccttttgaagatag---t 874  
 Qy 506 cgaccatgtgcattactcgtatttggccaatgctcctgctgattatttgaataatcagcag 565  
 Db 875 agaattatgcgaagtcatgttaattggcttaacttctccgaagaaaggaatcttttctga 934  
 Qy 566 ggacagagaataaaactg 581  
 Db 935 agatcgaataaaagtgtg 950

Search completed: October 24, 2001, 10:07:16  
 Job time: 621 sec

Db 611 agagtgagaaatcatcgccagccatgcatcgagagaaataatttgggtctatggccc 670  
 Qy 266 agacggagagagcggaagtccgtgtttgccaataatctcggactcaagcccgactgggtt 325  
 Db 671 aaatggagagagagaaacacgctatgcaaaacatctaatgaagacgagaaatgcgtt 730  
 Qy 326 ctacacatgtgtggagaccagaaagacgtattgtaccagtagcatcgagaccccaaacg 385  
 Db 731 ttattctccaggagaaatcattggatatatgtagactgtataattacaggatatat--- 787  
 Qy 386 aaatttaactcgtatgcacgggtgtaatttagaatttttaaaattatgcctgttaga 445  
 Db 788 ---tggtattatgtatattccaagatgcaaaagaggtatttttaaaattatgggttattaga 844  
 Qy 446 atgtgttaagaacacagggcattcagttcggaacaaatacgaaccccttagttatcttgggtt 505  
 Db 845 ggaatttaagaatggaataattcaaacggggaataatgaacccttttgaagatag---t 901  
 Qy 506 cgaccatgtgcattactcgtatttggccaatgctcctgctgattatttgaataatcagcag 565  
 Db 902 agaattatgcgaagtcatgttaattggcttaacttctccgaagaaaggaatcttttctga 961  
 Qy 566 ggacagagaataaaactg 581  
 Db 962 agatcgaataaaagtgtg 977

RESULT 15

AAV24084

ID AAV24084 standard; DNA; 1111 BP.

XX AC AAV24084;

XX DT 11-AUG-1998 (first entry)

XX DE Banana bunchy top virus component 1 DNA sequence.

XX KW BRTV; probe; diagnostic primer; component 1; ss.

XX OS Banana bunchy top virus.

XX PN US5756708-A.

XX PD 26-MAY-1998.

XX PF 24-FEB-1994; 94US-0202186.

XX PR 24-FEB-1994; 94US-0202186.

XX PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.

XX PI Burns TM, Dale JL, Harding RM, Karan M;

XX DR WPI; 1998-321636/28.

XX PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes

XX PT and primers and for producing virus-resistant plants

XX PS Disclosure; Column 17-18; 59pp; English.

XX CC This sequence represents a DNA fragment from the component 1 of

XX CC the Banana bunchy top virus (BRTV). The DNA molecules can be used as

XX CC diagnostic probes or primers or can be inserted into plants or other

XX CC organisms e.g. to produce virus-resistant plants or to act as promoters,

XX CC enhancers or termination signals.

XX SQ Sequence 1111 BP; 363 A; 192 C; 276 G; 280 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 7.2%; Score 59.6; DB 19; Length 1111;  
 Best Local Similarity 51.3%; Pred. No. 8.8e-10;  
 Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:40 ; Search time 5479.82 Seconds  
(without alignments)  
1436.950 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_383\_1215

Perfect score: 833

Sequence: 1 ggggtcttcgcagcagcga.....accatccagtgagtagct 833

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estomi:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*





```

/clone_lib="Barstead mouse myotubes MPLRB5"
/cell_line="C2C12"
/lab_host="DH10B"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTCAAGTCGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins. "
117 a 126 c 154 g 118 t 1 others

```

[illegible]

```

RESULT      4
AF0341173
LOCUS      2275 bp      mRNA      30-MAR-1998
DEFINITION      AF0341173 Human mRNA (Tripodis and Ragoussis) Homo sapiens CDNA
                clone ntcon2 contig, mRNA sequence.
ACCESSION      AF0341173
VERSION        AF0341173.1
KEYWORDS       GI:2707735
SOURCE        EST.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 2275)
AUTHORS        Tripodis,N. and Ragoussis,J.
TITLE          Generation of a transcription map in the region immediately
                centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
                boundary
JOURNAL        Unpublished (1997)
COMMENT        Contact: Tripodis, Nikos
                Division of Medical and Molecular Genetics
                Guys Hospital
                7th floor, Guy's Tower, London SE1 9RT, UK
                Email: nikos@ki.nl.
FEATURES       Location/Qualifiers
                1..2275
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /map="6p21.3"
                   /clone="ntcon2 contig"
                   /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a      619 c      470 g      599 t      149 others
ORIGIN
Query Match      4.6%; Score 38; DB 14; Length 2275;
Best Local Similarity 16.7%; Pred. NO. 1.1;
Matches 32; Conservative 85; Mismatches 75; Indels 0; Gaps

```

```

QY 570 agaataaaactgtgaataatttaaagtgtgtcatctaaataacacacacacccgcgc 629
Db 1469 AAAAAAAAAAAMWRYKXWKRGKRRKRTGMYKRYRAMMAMCMMACWYWKRGKKC 1528
QY 630 ccacgcgcgtactgtacatcttatgaatactctgccaggccgaaggcctggaggtg 689
Db 1529 WKYRYKTYSTYKSWRSRWYTYTYWYWCCTSMKSCAMRWGMYGSRSSRSY 1588
QY 690 ctaccgcgcgaagcgggaacaatatgaatcgatgtagtggcgccgcacataaaag 749
Db 1589 GYWGSMGCGYMKRYRYSWTGWTKWTWYWMSMTWTWYTYTWYTWRTTKTWYWW 1648
QY 750 attccattgaa 761
Db 1649 WTTCTWTKRGA 1660

RESULT 5
A0615572
LOCUS HS_5144_B1_B02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION HS_5144_B1_B02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION A0615572
VERSION A0615572.1 GI:5076848
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS 1 (bases 1 to 498)
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 720 row: D column: 3
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 498.
FEATURES
source
Location/Qualifiers
1..498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=720 Col=3 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 171 a 107 c 87 g 113 t 20 others
ORIGIN
Query Match 4.4%; Score 36.8; DB 231; Length 498;
Best Local Similarity 56.7%; Pred. No. 1.7;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 570 gtaagaacacagggcattcagttcggaacaaatacgaaccccttagttatcttggttcgac 509
Db 271 GTTCTGGCCAGGCAATCAGGCAAGACAACAATAAAGGGTATTAGTTAGGAAAAAGAG 330
QY 510 catgtcgatgactcgtatgttgcgaatgctgctgctgattattgaaatacagcaggagc 569
Db 331 GAAGTCAAAATGTCCCTGTTTGCAGATGACATGATTGTTTATTAGAAATCCCCATCGAC 390

RESULT 6
A0618171
LOCUS CM1-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.
DEFINITION CM1-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.
ACCESSION A0618171
VERSION A0618171.1 GI:7911165
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS 1 (bases 1 to 518)
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-cm1-ST0277-161
299-070-h05&t3=1999-12-16&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 76
High quality sequence stop: 183.
FEATURES
source
Location/Qualifiers
1..518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0277"
/dev_stage="Adult"
/note="organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 193 a 119 c 89 g 117 t
ORIGIN
Query Match 4.4%; Score 36.8; DB 121; Length 518;
Best Local Similarity 61.5%; Pred. No. 1.7;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 373 aggacccaaacgaatttaactcctgatgtaccagggtgatttagatttttaatt 432
Db 168 AGGATTAACCAAAATTTGATCTCATCGACACAGCAAAATATATAGTAACAGATT 227
QY 433 atgccctgttagaattgtttaagaacacagggcattca 468

```

```

512  tggcgtactcgtctattggccaatgtcctgctgattattgaaatacagcaggacag 571
Ddb  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
364  TATTATGACATAGAAATGGGAAAGCATTTCTCTATTAAATATGATCAGTAGCCATAT 305
Qy  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
572  aataaactgtggaatatttaaatgattgtgcatcctaataacaccaa 619
Ddb  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
304  GATATTGTTCTGAGGTGGAATCATTTTATTATTTAAATAACAAACA 257

RESULT 8
AQ089903/c
LOCUS
DEFINITION
  HS_3001_A2_H04_MR CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.
ACCESSION
  AQ089903
VERSION
  AQ089903.1 GI:3458814
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 318)
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  99380589
  MEDLINE
  COMMENT
    Contact: Mahairas GG, Wallace JC, Hood L
    High Throughput Sequencing Center
    University of Washington
    401 Queen Anne Avenue North, Seattle, WA 98109, USA
    Tel: (206) 616-3618
    Fax: (206) 616-3887
    Email: jwallace@u.washington.edu
    Sequence Tagged Connector
    Plate: 3001 row: O column: 8
    Class: BAC ends
    High quality sequence stop: 318.
    Location/Qualifiers
      1..318
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="Plate=3001 Col=8 Row=O"
        /clone.lib="CIT Approved Human Genomic Sperm Library D"
        /sex="male"
        /note="Organ: sperm; Vector: pBelorAC11; BAC Clones in
          E-Coli DH10B"
BASE COUNT 77 a 66 c 66 g 108 t 1 others
ORIGIN

Query Match 4.3%; Score 35.6; DB 224; Length 318;
Best Local Similarity 54.6%; Pred. No. 3.7;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 450 gtttaagaacaggcattcagttcggacaaatacgaaccccttagttatcttgggttcgac 509
Ddb 212 GTTCTGCCAGGCATCAGCGCAGGAGGAAGAAATAAAGGTTATTTAATTAGGAAGAGAG 153
Qy 510 catgtgcatgtactcgtattttgcgaatctcgtcgtgatttattgaaatacagcaggac 569
Ddb 152 GATTTCAAATTTGTCCTGTTTCAGATGACATGCTGTATATCTAGAAAACCGCATTTGC 93
Qy 570 agaataaaac 579
Ddb 92 TCAGTCCAAC 83

RESULT 9
CNS016H0
LOCUS
  CNS016H0
  11101 bp DNA GSS 26 JUL-1999

```

```

|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
228 GTGCTGAGAGTAACGTCTTACACTGCCCAATCA 263

RESULT      7
AZ444169/c
LOCUS       507 bp          DNA                      04-OCT-2000
DEFINITION  IM0239F20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0239F20 F, DNA sequence.
A2444169
A2444169.1 GI:10592981
GSS.
house mouse.
SOURCE      house mouse.
ORGANISM   Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 507)
Dunn,D., Aoyagi,A., Barber M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT,
84112 USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0239 row: F column: 20
Seq primer: CGTTGTAACAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 507.
Location/Qualifiers
    1..507
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="UUGC1M0239F20"
        /clone_lib="Mouse 10kb plasmid UUGC1M library"
        /sex="Male"
        /lab_host="E. Coli strain XL10-Cold, Tl-resistant, F-"
        /note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT  205 a     74 c     92 g     136 t
ORIGIN

```

Query Match 4.3%; Score 36; DB 244; Length 507;  
Best Local Similarity 58.3%; Pred. No. 3.1;  
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;



[illegible]

Thu Oct 25 13:08:06 2001

CENS0000X 488 bp DNA GSS 28-JUN-1999  
 LOCUS Arabidopsis thaliana genome survey sequence SP6 end of BAC F10F11  
 DEFINITION of IGF library from strain Columbia of Arabidopsis thaliana,  
 genomic survey sequence.  
 ACCESSION AL085215  
 VERSION AL085215.1 GI:5286355  
 KEYWORDS GSS.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicotids; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsids.  
 REFERENCE 1 (bases 1 to 488)  
 Salanoubat, M., Cholsne, N., Artiguenave, F., Brottier, P., Wincker, P.,  
 Samson, D., Saurin, W., Weissbach, J. and Quetier, F.  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 488)  
 REFERENCE Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr  
 FEATURES  
 source  
 1..488  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="IGF"  
 /clone="F10F11"  
 /note="end : SP6"  
 BASE COUNT 157 a 88 c 87 g 156 t  
 ORIGIN  
 Query Match 4.2%; Score 35.4; DB 219; Length 488;  
 Best Local Similarity 47.2%; Pred. No. 4.8;  
 Matches 108; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
 QY 447 tgggttaagaacaggcattcagtcggagacaataacgacaccccttagttatcttgggttc 506  
 |||||  
 Db 147 TCTTTTATCAACGACGGGTTTACCAAGAGAGAGAGAAATFACTAGATCATCTTGATTC 206  
 QY 507 gaccatgtcatgtactctgtatttcccaatgtcctgcctgattttgaaatcagcagg 566  
 |||||  
 Db 207 TTGTGTGACACTCTACATGATTTGAATTTTCCAAATTTACTTAACCTTAAGAAATGAGT 266  
 QY 567 gacagaataaacctgtggaattttaaagtatgtgtcatcttaataatcaccaataaccgc 626  
 |||||  
 Db 267 AATTGTGGAATTTAATTTGTTTTCAGGGTGTGTCAACTAGTACAAATCTATGCAACC 326  
 QY 627 ccgccaacgcgtatgattacatcttgaatctctgcccagccga 675  
 |||||  
 Db 327 TCGCTCTGGTGCAATCATCCCTCTTGAACACATACCAAGAGTCTGA 375  
 RESULT 12  
 BE603735 955 bp mRNA EST 02-MAR-2001  
 LOCUS HVSMEH0087003f Hordeum vulgare 5-45 DAP spike EST library  
 DEFINITION HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0087003f,  
 mRNA sequence.  
 ACCESSION BE603735  
 VERSION BE603735.2 GI:13187835  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare  
 ORGANISM Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
 ; Triticeae; Hordeum.  
 REFERENCE 1 (bases 1 to 955)  
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu  
 Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo  
 T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and

Wood, T.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics  
 JOURNAL Unpublished (2000)  
 COMMENT On Aug 21, 2000 this sequence version replaced gi:9861296.  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: AATTAACCTCACTAAAGGG  
 High quality sequence stop: 931.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..955  
 /organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone="HVSMEH0087003f"  
 /clone\_lib="Hordeum vulgare 5-45 DAP spike EST library"  
 HVCNDA0009 (5 to 45 DAP)  
 /tissue\_type="5-45 DAP Spike"  
 /lab\_host="SOLR"  
 /note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 303 a 205 c 213 g 220 t 14 others  
 ORIGIN  
 Query Match 4.2%; Score 35.4; DB 137; Length 955;  
 Best Local Similarity 52.0%; Pred. No. 5.7;  
 Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
 QY 268 acgagagagacgggaagtcctgtttgccaaatctctgagctcaagcccgactgttct 327  
 |||||  
 Db 585 ACGGGGACTGGAACCCCTGGGGTTACCAACCTAATCGGTTGACACATCCCTTTTC 644  
 QY 328 acacatgtgttggaacaggaagcgtattgtaccaggtacatcgaggagacccaaacgaa 387  
 |||||  
 Db 645 ACAGTTGGGTATAACAAAAAGGCGGACGATTCCCTCCAAAAGTGCCCAATCTGAA 704  
 QY 388 atttaactctgatgtaccaggtgaatt 417  
 |||||  
 Db 705 GGGGAATGGGAATGAGACGCTAATATT 734  
 RESULT 13  
 AZ401897/c 345 bp DNA GSS 03-OCT-2000  
 LOCUS IM0168N22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC1M0168N22 R, DNA sequence.  
 ACCESSION AZ401897  
 VERSION AZ401897.1 GI:10516971  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 345)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
 and Wright, D. Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177



|     | BASE COUNT | ORIGIN |
|-----|------------|--------|
| 1   | 0.000      | 0.000  |
| 2   | 0.000      | 0.000  |
| 3   | 0.000      | 0.000  |
| 4   | 0.000      | 0.000  |
| 5   | 0.000      | 0.000  |
| 6   | 0.000      | 0.000  |
| 7   | 0.000      | 0.000  |
| 8   | 0.000      | 0.000  |
| 9   | 0.000      | 0.000  |
| 10  | 0.000      | 0.000  |
| 11  | 0.000      | 0.000  |
| 12  | 0.000      | 0.000  |
| 13  | 0.000      | 0.000  |
| 14  | 0.000      | 0.000  |
| 15  | 0.000      | 0.000  |
| 16  | 0.000      | 0.000  |
| 17  | 0.000      | 0.000  |
| 18  | 0.000      | 0.000  |
| 19  | 0.000      | 0.000  |
| 20  | 0.000      | 0.000  |
| 21  | 0.000      | 0.000  |
| 22  | 0.000      | 0.000  |
| 23  | 0.000      | 0.000  |
| 24  | 0.000      | 0.000  |
| 25  | 0.000      | 0.000  |
| 26  | 0.000      | 0.000  |
| 27  | 0.000      | 0.000  |
| 28  | 0.000      | 0.000  |
| 29  | 0.000      | 0.000  |
| 30  | 0.000      | 0.000  |
| 31  | 0.000      | 0.000  |
| 32  | 0.000      | 0.000  |
| 33  | 0.000      | 0.000  |
| 34  | 0.000      | 0.000  |
| 35  | 0.000      | 0.000  |
| 36  | 0.000      | 0.000  |
| 37  | 0.000      | 0.000  |
| 38  | 0.000      | 0.000  |
| 39  | 0.000      | 0.000  |
| 40  | 0.000      | 0.000  |
| 41  | 0.000      | 0.000  |
| 42  | 0.000      | 0.000  |
| 43  | 0.000      | 0.000  |
| 44  | 0.000      | 0.000  |
| 45  | 0.000      | 0.000  |
| 46  | 0.000      | 0.000  |
| 47  | 0.000      | 0.000  |
| 48  | 0.000      | 0.000  |
| 49  | 0.000      | 0.000  |
| 50  | 0.000      | 0.000  |
| 51  | 0.000      | 0.000  |
| 52  | 0.000      | 0.000  |
| 53  | 0.000      | 0.000  |
| 54  | 0.000      | 0.000  |
| 55  | 0.000      | 0.000  |
| 56  | 0.000      | 0.000  |
| 57  | 0.000      | 0.000  |
| 58  | 0.000      | 0.000  |
| 59  | 0.000      | 0.000  |
| 60  | 0.000      | 0.000  |
| 61  | 0.000      | 0.000  |
| 62  | 0.000      | 0.000  |
| 63  | 0.000      | 0.000  |
| 64  | 0.000      | 0.000  |
| 65  | 0.000      | 0.000  |
| 66  | 0.000      | 0.000  |
| 67  | 0.000      | 0.000  |
| 68  | 0.000      | 0.000  |
| 69  | 0.000      | 0.000  |
| 70  | 0.000      | 0.000  |
| 71  | 0.000      | 0.000  |
| 72  | 0.000      | 0.000  |
| 73  | 0.000      | 0.000  |
| 74  | 0.000      | 0.000  |
| 75  | 0.000      | 0.000  |
| 76  | 0.000      | 0.000  |
| 77  | 0.000      | 0.000  |
| 78  | 0.000      | 0.000  |
| 79  | 0.000      | 0.000  |
| 80  | 0.000      | 0.000  |
| 81  | 0.000      | 0.000  |
| 82  | 0.000      | 0.000  |
| 83  | 0.000      | 0.000  |
| 84  | 0.000      | 0.000  |
| 85  | 0.000      | 0.000  |
| 86  | 0.000      | 0.000  |
| 87  | 0.000      | 0.000  |
| 88  | 0.000      | 0.000  |
| 89  | 0.000      | 0.000  |
| 90  | 0.000      | 0.000  |
| 91  | 0.000      | 0.000  |
| 92  | 0.000      | 0.000  |
| 93  | 0.000      | 0.000  |
| 94  | 0.000      | 0.000  |
| 95  | 0.000      | 0.000  |
| 96  | 0.000      | 0.000  |
| 97  | 0.000      | 0.000  |
| 98  | 0.000      | 0.000  |
| 99  | 0.000      | 0.000  |
| 100 | 0.000      | 0.000  |

|    | Query Match           | 4.2%  | Score 35.2;    | DB 233; | Length 848;    |
|----|-----------------------|---|----------------|---------|----------------|
|    | Best Local Similarity | 55.8%;  | Pred. No. 6.4; |         |                |
|    | Matches               | 67;   | Conservative   | 0;      | Mismatches 53; |
|    |                       |   |                |         | Indels 0;      |
|    |                       |   |                |         | Gaps 0;        |
| QY | 450                   | gttaagaacaggcgattcgaattcgacacatacgaaccccttagttatcttgggttcac | 509            |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |
| Db | 334                   | GTCTGGCCAGGCAATCAGGTAGGAGAAAGAAATAAAGCGTATTCATTTAGGAAAAGAG  | 393            |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |
| QY | 510                   | catgtgcgtactcgtatttggccaatgtcctgcctgattatttgaataatcagcggagc | 569            |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |
| Db | 394                   | GAAGTCAAATTTGCCCTGTTTGCAGATGACATGATTGTATTTAGAAAACCCCTTGTC   | 453            |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |
|    |                       |   |                |         |                |

Search completed: October 24, 2001, 13:13:43  
Job time: 11808 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:16 ; Search time 180.6 Seconds  
(without alignments)  
873.179 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_383\_1215  
Perfect score: 833  
Sequence: 1 gggtgtcttcgagcagcga.....accatccagtgagtagct 833

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA: \*  
1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2.6/ptodata/2/ina/PTUS\_COMB.seq: \*  
6: /cgn2.6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 120.6 | 14.5        | 1096   | 2     | US-08-418-071-4    |
| 2          | 117   | 14.0        | 1091   | 2     | US-08-418-071-5    |
| 3          | 102.2 | 12.3        | 1106   | 2     | US-08-418-071-3    |
| 4          | 99.2  | 11.9        | 1017   | 4     | US-08-793-634B-6   |
| 5          | 91.2  | 10.9        | 1022   | 4     | US-08-793-634B-2   |
| 6          | 62.8  | 7.5         | 1110   | 1     | US-08-202-186-14   |
| 7          | 59.6  | 7.2         | 982    | 3     | US-08-973-068-28   |
| 8          | 59.6  | 7.2         | 1110   | 1     | US-08-202-186-11   |
| 9          | 59.6  | 7.2         | 1111   | 1     | US-08-202-186-9    |
| 10         | 59.6  | 7.2         | 1111   | 1     | US-08-202-186-12   |
| 11         | 59.6  | 7.2         | 1111   | 1     | US-08-202-186-24   |
| 12         | 58.6  | 7.0         | 1103   | 1     | US-08-202-186-16   |
| 13         | 58.6  | 7.0         | 1105   | 1     | US-08-202-186-18   |
| 14         | 58    | 7.0         | 1109   | 1     | US-08-202-186-13   |
| 15         | 58    | 7.0         | 1111   | 1     | US-08-202-186-15   |
| 16         | 57    | 6.8         | 1104   | 1     | US-08-202-186-17   |
| 17         | 56.4  | 6.8         | 1111   | 1     | US-08-202-186-10   |
| 18         | 46.4  | 5.6         | 287    | 2     | US-08-418-071-1    |
| 19         | 46.4  | 5.6         | 300    | 2     | US-08-418-071-2    |
| 20         | 32.2  | 3.9         | 7218   | 1     | US-08-232-463-14   |
| 21         | 31    | 3.7         | 1212   | 3     | US-09-046-578-5    |
| 22         | 30.4  | 3.6         | 1212   | 3     | US-09-046-578-1    |
| 23         | 29.4  | 3.5         | 859    | 3     | US-09-280-409-2    |
| 24         | 29.4  | 3.5         | 1414   | 4     | US-08-686-968C-228 |
| 25         | 29.4  | 3.5         | 1437   | 3     | US-08-724-814-15   |
| 26         | 29.4  | 3.5         | 8920   | 2     | US-08-446-855A-1   |
| 27         | 29.4  | 3.5         | 8920   | 4     | US-09-150-741-1    |

|   |    |      |     |       |   |                   |                    |
|---|----|------|-----|-------|---|-------------------|--------------------|
| c | 28 | 29.2 | 3.5 | 1209  | 6 | 5352575-4         | Patent No. 5352575 |
|   | 29 | 29   | 3.5 | 9468  | 1 | US-08-325-547-10  | Sequence 10, Appl  |
|   | 30 | 28.8 | 3.5 | 1596  | 5 | PCT-US94-11328A-3 | Sequence 3, Appl   |
|   | c  | 31   | 3.5 | 10607 | 1 | US-08-078-090-3   | Sequence 3, Appl   |
|   | c  | 32   | 3.4 | 1593  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
|   | c  | 33   | 3.4 | 1593  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
|   | c  | 34   | 3.4 | 1593  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
|   | c  | 35   | 3.4 | 1593  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
|   | c  | 36   | 3.4 | 1608  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
|   | c  | 37   | 3.4 | 1608  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
|   | c  | 38   | 3.4 | 1608  | 2 | US-08-524-828-2   | Sequence 2, Appl   |
|   | c  | 39   | 3.4 | 2247  | 2 | US-08-975-114A-1  | Sequence 1, Appl   |
|   | c  | 40   | 3.4 | 2247  | 2 | US-08-975-114A-1  | Sequence 1, Appl   |
|   | c  | 41   | 3.4 | 2899  | 2 | US-08-624-581-2   | Sequence 1, Appl   |
|   | c  | 42   | 3.4 | 2917  | 2 | US-08-624-581-2   | Sequence 1, Appl   |
|   | c  | 43   | 3.4 | 2959  | 2 | US-08-624-581-2   | Sequence 1, Appl   |
|   | c  | 44   | 3.4 | 1380  | 4 | US-09-499-505-4   | Sequence 4, Appl   |
|   | c  | 45   | 3.4 | 1430  | 4 | US-09-499-505-3   | Sequence 3, Appl   |
|   | c  | 45   | 3.4 | 1556  | 2 | US-08-881-857-1   | Sequence 1, Appl   |

## ALIGNMENTS

RESULT 1  
US-08-418-071-4  
; Sequence 4, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rev-Yuh  
; APPLICANT: Soong, Tai-Seng  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/418.071  
FILING DATE: 06-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: DCB-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1096 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
MOLECULE TYPE: genomic DNA  
DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Banana Bunchy Top Virus (BBTV)  
US-08-418-071-4



|     | Query Match   | 12.3%;       | Score 102.2;     | DB 2;           | Length 1106;      |
|-----|---|--------------|------------------|-----------------|-------------------|
|     | Best Local Similarity   | 53.3%;       | Prod. NO. 5e-24; |                 |                   |
|     | Matches 215;  | Conservative | 0;               | Mismatches 188; | Indels 0; Gaps 0; |
| 195 | tgccagcttgaaagtgcgtgcgcatcggagagccagcgagcatcgcacaaatcctctgg     | 254          |                  |                 |                   |
| 524 | TGGCAGATTCAATTGACGGGAGCCAAATTGACGAGGAAACCCGATGATCGAGCATCATCTGG  | 583          |                  |                 |                   |
| 255 | atatcgagcagagcaggaagagcaggggagtcctgtgttgcacaaatctctcgactcaag    | 314          |                  |                 |                   |
| 584 | GTCTATGGTCTTATGGTAAATCAGGGTAAATCAACATATGGGAAAGTCACTAATCAAGAAG   | 643          |                  |                 |                   |
| 315 | cccgactggcttcacacatgtgtgtggaacgaaaggacgtatgtaccagtagcatatcgag   | 374          |                  |                 |                   |
| 644 | GATTGGTTCTACACCAAGGGGTGGGAAGAGGAGAAATATCTTAATTCCTATGTGGACGAA    | 703          |                  |                 |                   |
| 375 | gaccccaaacgaaatttaactcctcgtatcccgagggtgtaatttagagttattaaattat   | 434          |                  |                 |                   |
| 704 | GGATCTGCACACCATATAGTATTTTGATATTCCTCGTTGTAATCAGGATTAATTTAAATTAAT | 763          |                  |                 |                   |
| 435 | gcctcttgtagatgtgtttaagaacacgggcattcagttcggacaaataacgaccccttaqt  | 494          |                  |                 |                   |





|    |     |  |     |
|----|-----|--|-----|
| Db | 230 | ---tggttatatttgatatcccaagatgcaagaggattattttaattatgggtattatga   | 286 |
| QY | 446 | atgtgttaagaaacaggggcatttcagtttcgacaaatcgaacccctagttattcttgggtt | 505 |
| Db | 287 | ggatatttaagattggaataattcaaaagggaataatgaaccccttttgaagatag---t   | 343 |
| QY | 506 | cgaccatgctcatgtactcgtatttgcacatgtctcgtcgtattattgaaatcgcag      | 565 |
| Db | 344 | agaatattcgaaagtcattgtaatggctaaactctctccgaaggaaggaattcttctga    | 403 |
| QY | 566 | ggacagaataaaaactg  | 581 |
| Db | 404 | agatcgataaaattg  | 419 |

RESULT 8  
 US-08-202-186-11  
 : Sequence 11, Application US/08202186  
 : Patent No. 5756708  
 : GENERAL INFORMATION:  
 : APPLICANT: KARAN, Mirko  
 : APPLICANT: BURNS, Thomas M.  
 : APPLICANT: DALE, James L.  
 : APPLICANT: HARDING, Robert M.  
 : TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
 : NUMBER OF SEQUENCES: 60  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Foley & Lardner  
 : STREET: 3000 K Street, N.W.  
 : CITY: Washington, D.C.  
 : COUNTRY: USA  
 : ZIP: 20007-5109

STREET: 3000 K Street, N.W.  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,186  
FILING DATE: 24-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: JEFFERY, Donald D.  
REGISTRATION NUMBER: 19,980  
REFERENCE/DOCKET NUMBER: 71611/102 FIVE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1110 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
S-08-202-186-11

STRANDEDNESS: single  
 TOPOLOGY: circular  
 S-08-202-186-11

Query Match 7.2%; Score 59.6; DB 1; Length 1110;  
 Best Local Similarity 51.3%; Pred. No. 6.1e-10;  
 Matches 19; Conservative 0; Mismatches 174; Indels 9; Gaps

| Query Match           | 7.2%  | Score 59.6;        | DB 1;     | Length 1110; |
|-----------------------|---|--------------------|-----------|--------------|
| Best Local Similarity | 51.3%;  | Pred. NO. 6.1e-10; |           |              |
| Matches 193;          | Conservative 0;   | Mismatches 174;    | Indels 9; | Gaps         |
| 206                   | agtgtgtctgcgatacggagagccgagcgacatcgacacactctctggatatcgcgacg   | 265                |           |              |
| 584                   | AGAGGTGGAGAAATATCATGGCGAGCCATGTCATCGGAGAAATAATTGGGTCTATGGACC  | 643                |           |              |
| 266                   | agacggagagcgaggaagtcctgtttggcacaatatctcgactcaagccgcactgggt    | 325                |           |              |
| 644                   | AAATGGAGGAGAGGAAAGACACACCTATGC AAAACATCTAATGAAGACGAGAAATCCGTT | 703                |           |              |
| 326                   | ctacacatgtggtgggaaccagaaagcgattgtaccagtacatcgagaccaccaaacg    | 385                |           |              |
| 704                   | TTATTCTCCAGGAGGAAATCAATTGGATATGTAGACTCTATAATTAACGAGGATAT---   | 760                |           |              |

|   | Query Match           | 7.28;  | Score 59.6;        | DB 1;           | Length 1111;      |
|---|-----------------------|--|--------------------|-----------------|-------------------|
|   | Best Local Similarity | 52.6%;   | Pred. No. 6.1e-10; |                 |                   |
|   | Matches 181;          | Conservative   | 0;                 | Mismatches 154; | Indels 9; Gaps 2; |
| y | 238                   | atcgcacaaatcctcgtgatatcgagcagacgagagacggaagtcctgtttacca      | 297                |                 |                   |
|   |                       |  |                    |                 |                   |
| b | 616                   | ATTCGGAGATAATATTTGGGTCTATGGCCCAATATGGAGGAGAGAAACACACGTATGCCA | 675                |                 |                   |
|   |                       |  |                    |                 |                   |
| y | 298                   | atatctcggactcaagcccgactggttctcacacatggttggaaccagaaaggcgctat  | 357                |                 |                   |

[illegible]

```

Db 676 AACAACTAATGAGACGAGGAATGCTTTTATCTCCAGGGGAAATCAITGGATATAT 735
QY 358 tgtaccagtcacatcgagagaccacaaatatttaactcctcgatgacccagggtgaatt 417
Db 736 GTAGACTGTATATTAACGAGGNAT-----TGTATATTTGATATTCGAAGATGCAAG 789
QY 418 tagagtatttaaatatccctggttagaatgtgttaagaacaggcgatccagttcgaca 477
Db 790 AGGATTTATTAATTAATGCTTATTAGAGAAATTTAAGAAATGAATTAATCAAGCGGGA 849
QY 478 aatagcaaaccttagttatcttggttcgaccatgtgcagtactgtagtatttcccaatg 537
Db 850 AATATGAACCCGTTTGAAGATAG---TAGAATATGTGGAAGTCAATGTAATGCTAACT 906
QY 538 tcttcgctgattatttgaataatcgacgagggacagaaataaaactg 581
Db 907 TCCTTCCGAGGAAGGAATCTTTTCTGAGATCGAATAAAGTTG 950

RESULT 11
US-08-202-186-24
; Sequence 24, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 129...989
US-08-202-186-24

```

Query Match 7.2%; Score 59.6; DB 1; Length 1111;  
 Best Local Similarity 51.3%; Pred. No. 6.1e-10;  
 Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

```

QY 206 agtctgtctcgatcgagagccagcgagcagatcgcaaatcctctgtagatcgagcag 265
Db 611 AGAGGTGGAATATCATCGGCGCCATGTCTCATCGGAGAAATAATTTGGGTCTATGCCCC 670

```

```

QY 266 agacggaggagaggggagccgtgttgccaaatattctcgactcaagcccgactggtt 325
Db 671 AAATGGAGGAGAGGAAAGACAAAGTATGCAAAACATCTAATGAAGACGAGAAATCGGTT 730
QY 326 ctacacatgtggttggaaccagagaagagcgtattgtaccagtlacatcgagggacccaaacg 385
Db 731 TTATCTCCAGGAGGAAATCAATTTGGATATATGTAGACTGTATAATATACGAGGATAT--- 787
QY 386 aaatttaactcctgtagtaccaggtgtaattttagagatttttaaatatgcccctgttaga 445
Db 788 ---TGTTATATTTGATATTCCAAGATGCAAGAGGATTAATTTAAATATATGGGTATTAGA 844
QY 446 atgtgttaagaacaggcgatcagttcggacaaataacgaaccccttagttatcttgggtt 505
Db 845 GGAATTTAAGAATGGAATAATTCAAAGCGGAAATATGAACCGTTTGAAGATAG---T 901
QY 506 cgaccatgtgcagtactcgtatttgcgaatgtcctgcctgattatttgaataatcagcag 565
Db 902 AGAATATGCGAAGTCATTTGTAATGCTAACTTCCTTCGGAAGGAGGAATCTTTTCTGA 961
QY 566 ggacagaataaaactg 581
Db 962 AGATCGAATAAAGTTG 977

RESULT 12
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

```

Query Match 7.0%; Score 58.6; DB 1; Length 1103;  
 Best Local Similarity 52.5%; Pred. No. 1.3e-09;  
 Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;



Db 584 AGAGGTGGAGAAATATCATGCGCGACCATGTCATCGGAGAAATAATTTGGGTCTATGGCCC 643  
QY 266 agacggagagacaggaagtcggtgttggccaaatatctcgactcgaagcccgactgggt 325  
Db 644 AAATGGAGGAGGAAGAAAGCAACGTATGCAAAACGCTTAATGAAGACGAGAAATGGGTT 703  
QY 326 ctacacatgtgtggaacgaggaagtcggtgttggccaaatatctcgactcgaagcccgactgggt 325  
Db 704 TTATTTCCAGGAGGAGAAATCATTTGGATATATGTAGACTGTATAATTTACGAGGATAT--- 760  
QY 386 aaatttaactcctcgatgaccaggtgtaatttagagttatttaaatatgcccctgttaga 445  
Db 761 ---TGTTATATTTGATATTTCCAAAGATGCAAAACGCTTAATGAAGACGAGAAATGGGTTATTAGA 817  
QY 446 atgtgttaagaacagggcattcgtggaacgaggaagtcggtgttggccaaatatctcgactcgaagcccgactgggt 505  
Db 818 GGAATTTAAGAAATGGAATAATTTCAACCGGGAATATGAACCCGTTTGAAGATAG---T 874  
QY 506 cgaccatgtgcatgactcgtatttggccaaatatctcgactcgaagcccgactgggt 505  
Db 875 AGAATATGTCGAAGTCATTTGAATGCTAATTCCTTCCGAGGAAGGAATCTTTTCTGA 934  
QY 566 ggacagaataaaactg 581  
Db 935 AGATCGAATATAGTTG 950

RESULT 15  
US-08-202-186-15  
; Sequence 15, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JEFFERY, Donald D.  
; REGISTRATION NUMBER: 19,980  
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202 672 5300  
; TELEFAX: 202 672 5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1111 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
US-08-202-186-15

Query Match 7.0%; Score 58; DB 1; Length 1111;  
Best Local Similarity 51.1%; Pred. No. 2,1e-09;  
Matches 192; Conservative 0; Mismatches 175; Indels 9; Gaps 2;

QY 206 agtctgtctgcgatcgagagccagcgacatcgacaaatcctctggtatgacgagc 265  
Db 584 AGAGGTGGAGAAATATCATGCGCGACCATGTCATCGGAGAAATAATTTGGGTCTATGGCCC 643  
QY 266 agacggagagacaggaagtcggtgttggccaaatatctcgactcgaagcccgactgggt 325  
Db 644 AAATGGAGGAGGAAGAAAGCAACGTATGCAAAACGCTTAATGAAGACGAGAAATGGGTT 703  
QY 326 ctacacatgtgtggaacgaggaagtcggtgttggccaaatatctcgactcgaagcccgactgggt 325  
Db 704 TTATTTCCAGGAGGAGAAATCATTTGGATATATGTAGACTGTATAATTTACGAGGATAT--- 760  
QY 386 aaatttaactcctcgatgaccaggtgtaatttagagttatttaaatatgcccctgttaga 445  
Db 761 ---TGTTATATTTGATATTTCCAAAGATGCAAAACGCTTAATGAAGACGAGAAATGGGTTATTAGA 817  
QY 446 atgtgttaagaacagggcattcgtggaacgaggaagtcggtgttggccaaatatctcgactcgaagcccgactgggt 505  
Db 818 GGAATTTAAGAAATGGAATAATTTCAACCGGGAATATGAACCCGTTTGAAGATAG---T 874  
QY 506 cgaccatgtgcatgactcgtatttggccaaatatctcgactcgaagcccgactgggt 505  
Db 875 AGAATATGTCGAAGTCATTTGAATGCTAATTCCTTCCGAGGAAGGAATCTTTTCTGA 934  
QY 566 ggacagaataaaactg 581  
Db 935 AGATCGAATATAGTTG 950

Search completed: October 24, 2001, 10:00:19  
Job time: 204 sec

us-09-462-955-1\_copy\_383\_1215.rni

Thu Oct 25 13:08:04 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:43 ; Search time 5479.82 Seconds  
(without alignments)  
1347.248 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_211\_991  
Perfect score: 781  
Sequence: 1 gttgggacgaggttgacc.....ttaagtatgtgtcatctaa 781

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est33.\*  
26: gb\_est34.\*  
27: gb\_est35.\*  
28: gb\_est36.\*  
29: gb\_est37.\*  
30: gb\_est38.\*  
31: gb\_est39.\*  
32: gb\_est40.\*  
33: em\_estba.\*  
34: em\_estfun.\*  
35: em\_esthum1.\*  
36: em\_esthum2.\*  
37: em\_esthum3.\*  
38: em\_esthum4.\*  
39: em\_esthum5.\*  
40: em\_esthum6.\*  
41: em\_esthum7.\*  
42: em\_esthum8.\*  
43: em\_esthum9.\*

44: em\_esthum10.\*  
45: em\_esthum11.\*  
46: em\_esthum12.\*  
47: em\_esthum13.\*  
48: em\_esthum14.\*  
49: em\_esthum15.\*  
50: em\_esthum16.\*  
51: em\_esthum17.\*  
52: em\_esthum18.\*  
53: em\_esthum19.\*  
54: em\_esthum20.\*  
55: em\_esthum21.\*  
56: em\_esthum22.\*  
57: em\_esthum23.\*  
58: em\_esthum24.\*  
59: em\_esthum25.\*  
60: em\_esthum26.\*  
61: em\_esthum27.\*  
62: em\_esthum28.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estin5.\*  
68: em\_estom1.\*  
69: em\_estom2.\*  
70: em\_estov1.\*  
71: em\_estov2.\*  
72: em\_estpl1.\*  
73: em\_estpl2.\*  
74: em\_estpl3.\*  
75: em\_estpl4.\*  
76: em\_estpl5.\*  
77: em\_estpl6.\*  
78: em\_estpl7.\*  
79: em\_estpl8.\*  
80: em\_estpl9.\*  
81: em\_estpl10.\*  
82: em\_estrol.\*  
83: em\_estrol2.\*  
84: em\_estrol3.\*  
85: em\_estrol4.\*  
86: em\_estrol5.\*  
87: em\_estrol6.\*  
88: em\_estrol7.\*  
89: em\_estrol8.\*  
90: em\_estrol9.\*  
91: em\_estrol10.\*  
92: em\_estrol11.\*  
93: em\_estrol12.\*  
94: em\_estrol13.\*  
95: em\_estrol14.\*  
96: em\_estrol15.\*  
97: em\_estrol16.\*  
98: em\_estrol17.\*  
99: em\_estrol18.\*  
100: em\_estrol19.\*  
101: em\_estrol20.\*  
102: gb\_est25.\*  
103: gb\_est26.\*  
104: gb\_est27.\*  
105: gb\_est28.\*  
106: gb\_est29.\*  
107: gb\_est30.\*  
108: gb\_est31.\*  
109: gb\_est32.\*  
110: gb\_est41.\*  
111: gb\_est42.\*  
112: gb\_est43.\*  
113: gb\_est44.\*  
114: gb\_est45.\*  
115: gb\_est46.\*  
116: gb\_est47.\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.





```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:594576"
/clone_lib="Stratagene ovarian cancer (#937219)"
/sex="female"
/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
/Note="vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Papillary
serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'
BASE COUNT 130 a 56 c 101 g 120 t 6 others
ORIGIN

```

```

Query Match 4.9%; Score 38.2; DB 3; Length 413;
Best Local Similarity 54.7%; Pred. NO. 0.58;
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 273 ccagcgccgacccgaagatgcgttggtacacgagccttcggtggaatggacgaagatggc 332
DB 140 CTCGTGATGTCAGAGAGAGAGTCTTTAGGAAGCTTCACTTTAAAGGACAAAGCGA 199

QY 333 cgcgtgaaatccgttccattccatatacacaaattggcagcttgagtgctgcgcat 392
DB 200 TTTGAAAAAGCCCTAACACAGGATTCAGGTCCTCGGTTTGAGTCTTGAAGAATTTATTGCTTT 259

QY 393 cggagagccagcgagcat 411
DB 260 TGAGCATCTCGAAGAAGTT 278

```

```

RESULT 3
AA516896 516 bp mRNA EST 14-JUL-1997
LOCUS v979e09.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
DEFINITION IMAGE:902248.5; similar to SW:RS9_HUMAN P46781 40S RIBOSOMAL
PROTEIN S9. ; mRNA sequence.
ACCESSION AA516896
VERSION AA516896
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 516)
REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
AUTHORS Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:529212
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 473.
Location/Qualifiers
1. .516
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:902248"
FEATURES
source

```

```

/clone_lib="Barstead mouse myotubes MPLRB5"
/cell_line="C2C12"
/lab_host="DH10B"
/Note="Vector: pWT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACGAATCTGAAGTGGAGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins.
BASE COUNT 117 a 126 c 154 g 118 t 1 others
ORIGIN

```

```

Query Match 4.9%; Score 38.2; DB 8; Length 516;
Best Local Similarity 51.8%; Pred. NO. 0.62;
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 237 atttgctggaacacctgataactccgctggaagccagcgatgccgaagatgcgt 296
DB 38 ATTGTCTAGCANAACACTATGTGAACCCACGGAGAACCTTCGGAGAAAGTTGGCGTCT 97

QY 297 tgtaacacgagcttcggtggaatggacaaatggccgctgaaatccgttccattcc 356
DB 98 TGGACAGGAGCTAAAGTTGATTGGAGAGTATGACTCCGGAACAAACGTTGAGTTTGA 157

QY 357 atatcaaatggcagcttgaaatgctgctgcatcggaagac 400
DB 158 GGGTCAAGTTTACCTCGCCAAGATCGTAAGCGGCCCGGAGC 201

```

```

RESULT 4
AA615572 498 bp DNA GSS 15-JUN-1999
LOCUS HS_5144_B1_B02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=720 Col=3 Row=D, DNA sequence.
ACCESSION AA615572
VERSION AA615572.1 GI:5076848
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 498)
REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
AUTHORS Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.P. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Libraries are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieteredejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
plate: 720 row: D column: 3
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 498.
Location/Qualifiers

```

```

source
1. .498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=720 Col=3 Row=D"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT      171 a   107 c   87 g   113 t   20 others
ORIGIN

Query Match      4.7%; Score 36.8; DB 231; Length 498;
Best Local Similarity 56.7%; Pred. No. 1.7;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 622 gtaagaacaggcattcagtcgacaaataacgaccccttagtattcttggttcgac 681
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 GTTCTGGCCAGGCAATCAGGCAAGACAAAGAAATAAGGGTATTAACTTAGGAAAAGAG 330
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 682 catgtgatgtactgtatttgcacatgctcctgctgatttggaaatcagcaggac 741
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 GAAGTCAAAATGTCCCTGTTTGCAGATGACATGATGTTTATTATAGAAATCCCATCGAC 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AW818171      518 bp      mRNA      EST      17-MAY-2000
LOCUS      CMI-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.
ACCESSION      AW818171
VERSION      AW818171.1 GI:7911165
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 518)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CMI-ST0277-161
299-070-h05&t3=1999-12-16&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 76
High quality sequence stop: 183.
Location/Qualifiers
1. .518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0277"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
source

1. .498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=720 Col=3 Row=D"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT      171 a   107 c   87 g   113 t   20 others
ORIGIN

Query Match      4.7%; Score 36.8; DB 231; Length 498;
Best Local Similarity 56.7%; Pred. No. 1.7;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 622 gtaagaacaggcattcagtcgacaaataacgaccccttagtattcttggttcgac 681
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 GTTCTGGCCAGGCAATCAGGCAAGACAAAGAAATAAGGGTATTAACTTAGGAAAAGAG 330
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 682 catgtgatgtactgtatttgcacatgctcctgctgatttggaaatcagcaggac 741
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 GAAGTCAAAATGTCCCTGTTTGCAGATGACATGATGTTTATTATAGAAATCCCATCGAC 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AW818171      518 bp      mRNA      EST      17-MAY-2000
LOCUS      CMI-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.
ACCESSION      AW818171
VERSION      AW818171.1 GI:7911165
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 518)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CMI-ST0277-161
299-070-h05&t3=1999-12-16&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 76
High quality sequence stop: 183.
Location/Qualifiers
1. .518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0277"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
source

derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      193 a   119 c   89 g   117 t
ORIGIN

Query Match      4.7%; Score 36.8; DB 121; Length 518;
Best Local Similarity 61.5%; Pred. No. 1.7;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 545 aggaccccaaacgaaatttaactcgtatgccacagggtgaatttagatttaatt 604
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 AGGATTAAACCAATTTTCATCTCACATGGACACAGCAAGCAAAAATATAGTACAAAGATT 227
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 605 atgccctgttagtaattgtgttaagaacagggtcatcca 640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 228 GTGCTCAGAAAGTAACGTCTTAACACCTGCCCAATCA 263
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
AQ089903      318 bp      DNA      GSS      26-AUG-1998
LOCUS      HS_3001_A2_H04_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.
ACCESSION      AQ089903
VERSION      AQ089903.1 GI:3458814
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 318)
Mabairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mabairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3001 row: O column: 8
Class: BAC ends
High quality sequence stop: 318.
Location/Qualifiers
1. .318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3001 Col=8 Row=O"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      77 a   66 c   66 g   108 t   1 others
ORIGIN

Query Match      4.6%; Score 35.6; DB 224; Length 318;
Best Local Similarity 54.6%; Pred. No. 3.6;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 622 gtaagaacaggcattcagtcgacaaataacgaccccttagtattcttggttcgac 681
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 GTTCTGCCAGGCAATCAGGCAAGCAAGCAAAAATAGGTTATTTAATTAGGAAGAG 153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

were sequenced additional times to assess quality

```

QY 682 catgtgcatgtactgtatttgcacaaatgctgctgctgattatttgaataatcagcaggag 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 GATTTCAAAATGTCCTCTGTTGCAGATGACATGACTGTATATCTAGAAACCGCATGTC 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 742 agataaaac 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 TCAGTCCAAC 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
A1942816 460 bp mRNA EST 06-AUG-1999
LOCUS f657b05.xl zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA
DEFINITION sequence.
ACCESSION A1942816
VERSION A1942816.1 GI:5707472
KEYWORDS zebrafish.
SOURCE EST.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 460)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
WashU zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResearchGeneticsPrimerBank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T7 ET from Amersham
High quality sequence stop: 443.
FEATURES
source
1..460
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish washU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'-pGACTAGTCTAGATCCGAGCGCGCCCTTTTCTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones

```

were sequenced additional times to assess quality  
 control." 154 a 60 c 59 g 187 t  
 BASE COUNT 154 a 60 c 59 g 187 t  
 ORIGIN  
 Query Match 4.5%; Score 35.4; DB 104; Length 460;  
 Best Local Similarity 50.9%; Pred. No. 4.6;  
 Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
 QY 595 tatttaaatatgcctgttagaattgttaagaacagggcatttcgacgaataac 654  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 5 TTTTITTTTTTACAAATTTGAAGTGTGTTTACAAAAACAATTAATTTGTCCAAAAA 64  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 QY 655 gaaccccttagttatcttgggttcgaccatgtgcattgtactctgtatttgcacatgcctg 714  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 65 AACCTCCATAAATATATGTTGTTTAAAGCTCTATTATAAATCAGTAGTTTGAACAACAACCAAG 124  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 QY 715 cctgattatttgaataatcagcaggacagagaataaaactgtggaat 759  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 125 TATAATTTTGTGAATCACTAGTGAATTAATATTATTCATGTCGCAT 169  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 RESULT 8  
 BE603735 955 bp mRNA EST 02-MAR-2001  
 LOCUS HVSMEH0087003f Hordeum vulgare 5-45 DAP spike EST library  
 DEFINITION HVCDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0087003f,  
 mRNA sequence.  
 ACCESSION BE603735  
 VERSION BE603735.2 GI:13187835  
 KEYWORDS EST.  
 SOURCE barley.  
 ORGANISM Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
 ; Triticeae; Hordeum.  
 REFERENCE 1 (bases 1 to 955)  
 AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu  
 ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo  
 ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D., and  
 Wood,T.  
 TITLE Development of a genetically and physically anchored EST resource  
 for barley genomics  
 JOURNAL Unpublished (2000)  
 COMMENT On Aug 21, 2000 this sequence version replaced gi:9861296.  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: AATTAACCCCTCACTAAAGG  
 High quality sequence stop: 931.  
 FEATURES  
 source  
 1..955  
 /organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone\_lib="Hordeum vulgare 5-45 DAP spike EST library  
 HVCDNA0009 (5 to 45 DAP)"  
 /tissue\_type="5-45 DAP Spike"  
 /lab\_host="SOLR"  
 /note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 303 a 205 c 213 g 220 t 14 others  
 ORIGIN  
 Query Match 4.5%; Score 35.4; DB 137; Length 955;  
 Best Local Similarity 52.0%; Pred. No. 5.7;  
 Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;



Thu Oct 25 13:08:00 2001

```

Qy 459 cgtgtttccaaatctcggactcaagccgactgttc 498
Db 754 TTAGATGGCAAGCGTCCAGATGAAGTCCGCCGGTTC 793

RESULT 11
LOCUS      AQ746046      848 bp      DNA      GSS      16-JUL-1999
DEFINITION HS_2277_A2_F11_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2277 Col=22 Row=K, DNA sequence.
ACCESSION  AQ746046
VERSION     AQ746046.1 GI:523568
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            I (bases 1 to 848)
REFERENCE  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
AUTHORS    Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
            Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2277 row: K column: 22
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 848.
            Location/Qualifiers
            1..848
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate=2277 Col=22 Row=K"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
            E-Coli Dh10B"
BASE COUNT 228 a 136 c 117 g 144 t 223 others
ORIGIN

Query Match 4.5%; Score 35.2; DB 233; Length 848;
Best Local Similarity 55.8%; Pred. No. 6.4;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 622 gttagaacagggcattcagtcggacaaatacgaaccccttagttatcttgggttcgac 681
Db 334 GTTCTGCCAGGGCAATCAGGTAGGAGAAAGAAATAAACGGTATTCAATTAGGAAAAGAG 393

Qy 682 catgtgcattcgtctattgccaatgtcctgcctgattattgaaaatacagcaggac 741
Db 394 GAAGTCAAAATGTCCCTGTTTGCAGATGACATGATTGTATATTAGAAAACCCCATTTGC 453

RESULT 12
LOCUS      AQ418525      487 bp      DNA      GSS      23-MAR-1999
DEFINITION RPCI-11-20318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-20318,
            DNA sequence.
ACCESSION  AQ418525
VERSION     AQ418525.1 GI:4476249
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            I (bases 1 to 487)
REFERENCE  Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter
AUTHORS    J.C.
TITLE      Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Qy 608 cctctttagaatgtttaagaacagggcattcagtcggacaaatacgaaccccttagtt 667
Db 127 CTCCTGGTAAATTCGGTCCGTCAGGGCAATCAGGCGAGGAGAAAGAAATAAAGGGTGTTC 186

Qy 668 atcttggttcgacatgtcattctactcgtatttgcgaatgtcctgcctgattattga 727
Db 187 AATTAGGAAAAGAGGAAGTCAATGTCCCTGTTTGCAGATGACATGATTGTGTATCTAC 246

Qy 728 aaatcagca 736
Db 247 AAAACCCCA 255

Query Match 4.4%; Score 34.6; DB 228; Length 487;
Best Local Similarity 54.3%; Pred. No. 8.4;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 608 cctctttagaatgtttaagaacagggcattcagtcggacaaatacgaaccccttagtt 667
Db 127 CTCCTGGTAAATTCGGTCCGTCAGGGCAATCAGGCGAGGAGAAAGAAATAAAGGGTGTTC 186

Qy 668 atcttggttcgacatgtcattctactcgtatttgcgaatgtcctgcctgattattga 727
Db 187 AATTAGGAAAAGAGGAAGTCAATGTCCCTGTTTGCAGATGACATGATTGTGTATCTAC 246

Qy 728 aaatcagca 736
Db 247 AAAACCCCA 255

RESULT 13
LOCUS      AQ417598      555 bp      DNA      GSS      23-MAR-1999
DEFINITION RPCI-11-203C8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203C8,
            DNA sequence.
ACCESSION  AQ417598
VERSION     AQ417598.1 GI:4475441
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            I (bases 1 to 555)
REFERENCE  Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter
AUTHORS    J.C.
TITLE      Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

```

**JOURNAL COMMENT**

Map Building  
 Unpublished (1997)  
 Other\_GSSs: RPCI-11-203C8.TJ  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([inforesgen.com](http://inforesgen.com)). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
 Seq primer: T7  
 Class: BAC ends.

**FEATURES** source Location/Qualifiers

1..555  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7577623"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-203C8"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="lymphocytes"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPc11 Human Male BAC Library"

**BASE COUNT** 192 a 114 c 110 g 139 t

**ORIGIN**

Query Match 4.4%; Score 34.6; DB 228; Length 555;  
 Best Local Similarity 54.3%; Pred. No. 8.7;  
 Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 608 cccgttagaattgttaaacagagcattcagtcgacaaataacgaccccttagtt 667  
 Db 150 CTCGTGTAGATTGCGTCCGTCAGGCAATCATCAGGAGAGAAATAAAGGGTTC 209  
 Qy 668 atctgttgaccatgtgcattgctgatttcccaatgtcctgctgatttga 727  
 Db 210 AATTAGAAAGAGGAAGTCAAAATTCCTCTTTCACATGACATGATGTATCTAC 269  
 Qy 728 aaatcagca 736  
 Db 270 AAAACCCCA 278

**RESULT** 14  
 CNS016H0  
 LOCUS  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN16B20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL106734  
 VERSION AL106734.1 GI:5623558  
 KEYWORDS GSS  
 SOURCE fruit fly.  
 ORGANISM Plasmid Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1101)  
 AUTHORS Direct Submission  
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY.cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))  
 JOURNAL - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

**FEATURES** source Location/Qualifiers

1..1101  
 /organism="Drosophila melanogaster"  
 /plasmid="pBelOBAC11"  
 /db\_xref="taxon:7227"  
 /clone\_lib="DrosBAC"  
 /clone="BACN16B30"  
 /note="end : T7"

**BASE COUNT** 260 a 123 c 159 g 217 t 342 others

**ORIGIN**

Query Match 4.4%; Score 34.6; DB 219; Length 1101;  
 Best Local Similarity 30.8%; Pred. No. 11;  
 Matches 57; Conservative 47; Mismatches 81; Indels 0; Gaps 0;

Qy 590 tagagattttaattatgcccttgatgaattgtttaaacagagcattcagtcgaca 649  
 Db 666 TATATWADTWATWATWATTAATGTTTTRMASAAARPTHKGGAKADWTATNTRRGR 725  
 Qy 650 aatacgaaccccttagttatcttgggttcacccatgtgcattgctgatttgcattg 709  
 Db 726 DGGCGAACAGYKATGATGATGKAKRGAARGGKWTGTRTARTATWGGDRKAGW 785  
 Qy 710 tctgcctgatttgaataacagcagagacagataaaactgtggaattattaaagta 769  
 Db 786 TTGKTKTKTKAWCKTDKAKTGARTGAGGGAGBAGAVTGWATWTDABTSDMRGAG 845  
 Qy 770 tgtgt 774  
 Db 846 NRTAT 850

**RESULT** 15  
 BGL14850/c  
 LOCUS BGL14850 788 bp mRNA EST 30-JAN-2001  
 DEFINITION 602315618F1 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:4415933 5', mRNA sequence.  
 ACCESSION BGL14850  
 VERSION BGL14850.1 GI:12608356  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 788)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 Cloning by: Incyte Genomics, Inc.  
 Cloning distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLAM10146 row: c column: 06  
 High quality sequence stop: 733.  
 Location/Qualifiers  
 1..788  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4415933"  
 /clone\_lib="NIH\_MGC\_88"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"

**FEATURES** source Location/Qualifiers

1..788  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4415933"  
 /clone\_lib="NIH\_MGC\_88"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"

Thu Oct 25 13:08:00 2001

/note="Organ: small intestine; Vector: pCMV-SPORT6;  
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;  
oligo-dT primed. Average insert size 1.767 kb. Library  
enriched for full-length clones and constructed by Life  
Technologies. Note: this is a NIH\_MGC Library."  
BASE COUNT 126 a 260 c 245 g 157 t  
ORIGIN

|                       |                 |                |           |             |
|-----------------------|-----------------|----------------|-----------|-------------|
| Query Match           | 4.4%            | Score 34.4;    | DB 174;   | Length 788; |
| Best Local Similarity | 52.0%;          | Pred. No. 11;  |           |             |
| Matches 77;           | Conservative 0; | Mismatches 71; | Indels 0; | Gaps 0;     |

|    |     |   |     |
|----|-----|---|-----|
| QY | 55  | cacttgaagacccggtcggcgactgcaaggattgaagactgttcttgggaatgacaggatt | 114 |
| DB | 566 | CCCAGGAGACAGAGGCTGGACAGCATGGAGCTGCAGGTGATCAGTCAATGACATTGTC    | 507 |
| QY | 115 | cacctggagccgacccgtgtgttcgcagcaacaagaatagagactactgttcgaaggacgg | 174 |
| DB | 506 | CAGCTGCTGACGACCGGACCGCCGGCCACCAAGTGCACCGGCTCCAGCAGGAGATGAC    | 447 |
| QY | 175 | gtcttctcgagcacggagtcctccgactc                                 | 202 |
| DB | 446 | GGCCGTCTCCAGCACGTTGCTCCCGCTC                                  | 419 |

Search completed: October 24, 2001, 13:13:45  
Job time: 11810 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:20 ; Search time 393.52 Seconds  
(without alignments)  
930.236 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_409\_991  
Perfect score: 583  
Sequence: I actgctcctggagtcacaaag.....ttaaagtatgtgcatctaa 583

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 116.4 | 20.0        | 1096   | 20 | AAV71833    |
| 2          | 116.4 | 20.0        | 1096   | 21 | AAV71833    |
| 3          | 114.8 | 19.7        | 1091   | 20 | AAV71834    |
| 4          | 114.8 | 19.7        | 1091   | 21 | AAV71834    |
| 5          | 114.4 | 19.6        | 886    | 20 | AAV71832    |
| 6          | 102.2 | 17.5        | 1106   | 20 | AAV71832    |
| 7          | 100.6 | 17.3        | 1106   | 21 | AAV71832    |
| 8          | 99.2  | 17.0        | 1017   | 17 | AAV713165   |
| 9          | 91.2  | 15.6        | 1022   | 17 | AAV713161   |
| 10         | 80.4  | 13.8        | 593    | 20 | AAV71832    |
| 11         | 62.8  | 10.8        | 1110   | 19 | AAV24089    |

|    |      |      |       |    |          |                    |
|----|------|------|-------|----|----------|--------------------|
| 12 | 59.6 | 10.2 | 982   | 18 | AAV71833 | Banana bunchy top  |
| 13 | 59.6 | 10.2 | 1110  | 19 | AAV24086 | Banana bunchy top  |
| 14 | 59.6 | 10.2 | 1111  | 19 | AAV24077 | Banana bunchy top  |
| 15 | 59.6 | 10.2 | 1111  | 19 | AAV24084 | Banana bunchy top  |
| 16 | 59.6 | 10.2 | 1111  | 19 | AAV24087 | Banana bunchy top  |
| 17 | 58.6 | 10.1 | 1103  | 19 | AAV24091 | Banana bunchy top  |
| 18 | 58.6 | 10.1 | 1105  | 19 | AAV24093 | Banana bunchy top  |
| 19 | 58   | 9.9  | 1109  | 19 | AAV24088 | Banana bunchy top  |
| 20 | 58   | 9.9  | 1111  | 19 | AAV24090 | Banana bunchy top  |
| 21 | 57   | 9.8  | 1104  | 19 | AAV24092 | Banana bunchy top  |
| 22 | 56.4 | 9.7  | 1111  | 19 | AAV24085 | Banana bunchy top  |
| 23 | 46.4 | 8.0  | 287   | 20 | AAV71830 | Subgenomic fragmen |
| 24 | 46.4 | 8.0  | 287   | 21 | AAV38948 | Banana bunchy top  |
| 25 | 45.8 | 7.9  | 936   | 22 | AAV58252 | Oligonucleotide D1 |
| 26 | 45.8 | 7.9  | 936   | 22 | AAV58254 | Oligonucleotide D1 |
| 27 | 45.8 | 7.9  | 936   | 22 | AAV58257 | Oligonucleotide D1 |
| 28 | 45.8 | 7.9  | 936   | 22 | AAV58259 | Oligonucleotide D2 |
| 29 | 45.8 | 7.9  | 936   | 22 | AAV58262 | Oligonucleotide D2 |
| 30 | 45.8 | 7.9  | 938   | 22 | AAV58255 | Oligonucleotide D1 |
| 31 | 44.2 | 7.6  | 300   | 20 | AAV71831 | Subgenomic fragmen |
| 32 | 44.2 | 7.6  | 300   | 21 | AAV38949 | Banana bunchy top  |
| 33 | 42.4 | 7.3  | 936   | 22 | AAV58252 | Oligonucleotide D1 |
| 34 | 42.4 | 7.3  | 936   | 22 | AAV58254 | Oligonucleotide D1 |
| 35 | 42.4 | 7.3  | 936   | 22 | AAV58257 | Oligonucleotide D1 |
| 36 | 42.4 | 7.3  | 936   | 22 | AAV58259 | Oligonucleotide D1 |
| 37 | 42.4 | 7.3  | 936   | 22 | AAV58262 | Oligonucleotide D2 |
| 38 | 42.4 | 7.3  | 936   | 22 | AAV58255 | Oligonucleotide D1 |
| 39 | 37   | 6.3  | 244   | 22 | AAV58238 | Oligonucleotide D1 |
| 40 | 34.2 | 5.9  | 244   | 22 | AAV58238 | Oligonucleotide D1 |
| 41 | 32.6 | 5.6  | 5053  | 19 | AAV22816 | Nucleotide sequenc |
| 42 | 32.6 | 5.6  | 5055  | 19 | AAV17236 | DNA from a region  |
| 43 | 31.4 | 5.4  | 41599 | 21 | AAZ35351 | Cosmid including s |
| 44 | 31   | 5.3  | 1212  | 20 | AAV06636 | Bacillus anthracis |
| 45 | 31   | 5.3  | 1212  | 20 | AAV06645 | Bacillus anthracis |

ALIGNMENTS

RESULT 1  
AAV71833  
ID AAV71833 standard; DNA; 1096 BP.  
XX  
AC AAV71833;  
XX  
DT 10-FEB-1999 (first entry)  
XX  
DE BBTV DNA II clone (2-17) nucleotide sequence.  
XX  
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
XX  
OS Banana bunchy top virus.  
XX  
FH Key Location/Qualifiers  
FT TATA\_signal 1..7  
FT /\*tag= a  
FT stem\_loop 8..38  
FT /\*tag= b  
FT CDS 70..928  
FT /\*tag= c  
FT /\*product= "ORF-V2 product"  
FT /\*transl\_except= (pos:215..217, aa:Gly)  
FT polyA\_signal 533..538  
FT /\*tag= d  
FT polyA\_signal 795..804  
FT /\*tag= e  
FT polyA\_signal 907..912  
FT /\*tag= f  
FT polyA\_signal 1030..1035  
FT /\*tag= ge  
XX  
PN US5846705-A.





|                          |                                  |
|--------------------------|----------------------------------|
| RESULT                   | 6                                |
| AAV71832                 |                                  |
| ID                       | AAV71832 standard; DNA; 1106 BP. |
| XX                       |                                  |
| XX                       |                                  |
| AAV71832;                |                                  |
| AC                       |                                  |
| DT                       |                                  |
| 10-FEB-1999              | (first entry)                    |
| RTTV DNA I clone (7-4-2) | nucleotide sequence.             |
| DE                       |                                  |
| XX                       |                                  |







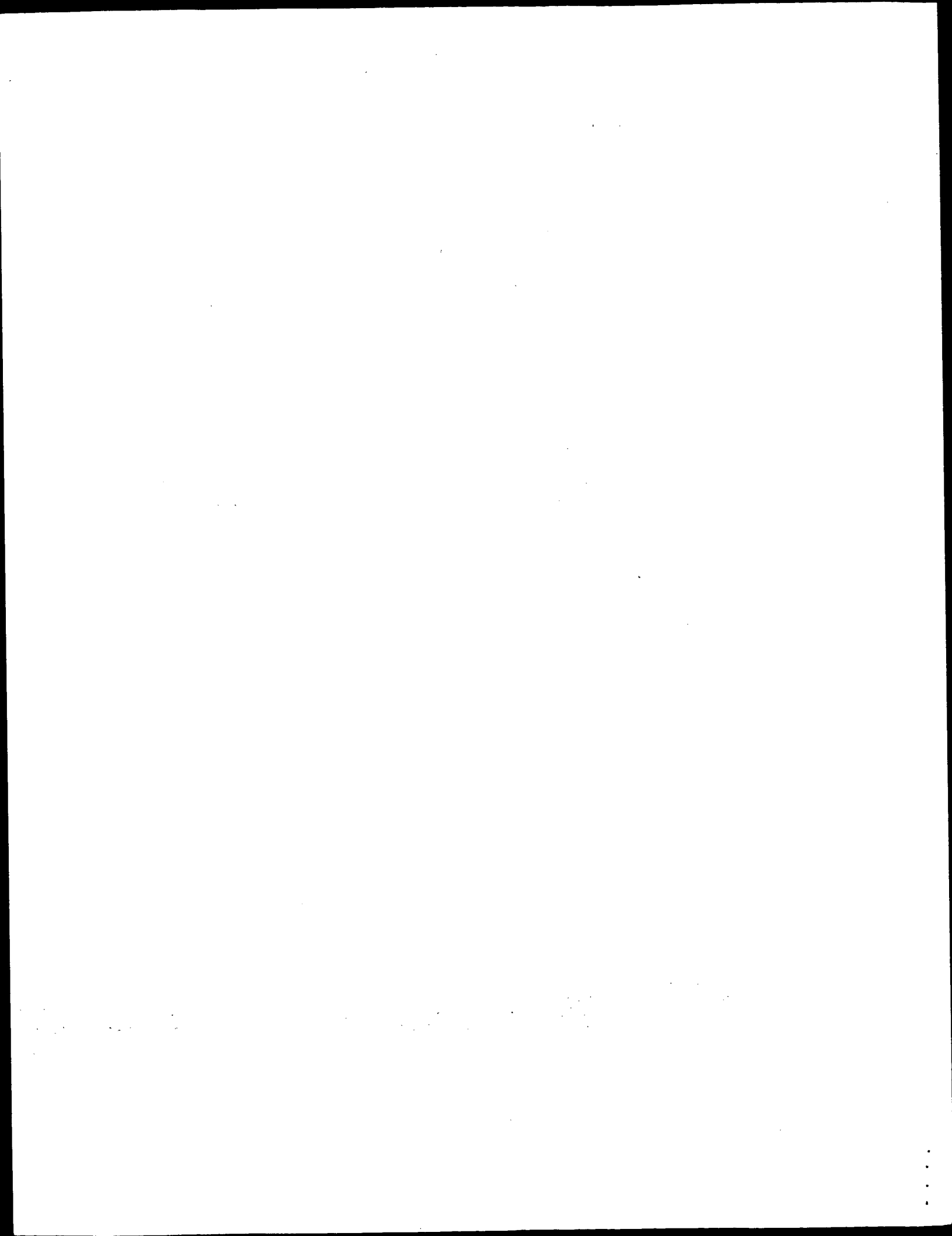












GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:45 ; Search time 5479.82 seconds  
(without alignments)  
1005.692 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_409\_991

Perfect score: 583

Sequence: 1 actgcctggagtgcaaaag.....ttaaagtatgtgcataa 583

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

|      |               |
|------|---------------|
| 1:   | gb_estl1:*    |
| 2:   | gb_est2:*     |
| 3:   | gb_est3:*     |
| 4:   | gb_est4:*     |
| 5:   | gb_est5:*     |
| 6:   | gb_est6:*     |
| 7:   | gb_est7:*     |
| 8:   | gb_est8:*     |
| 9:   | gb_est9:*     |
| 10:  | gb_est10:*    |
| 11:  | gb_est11:*    |
| 12:  | gb_est12:*    |
| 13:  | gb_est13:*    |
| 14:  | gb_est14:*    |
| 15:  | gb_est15:*    |
| 16:  | gb_est16:*    |
| 17:  | gb_est17:*    |
| 18:  | gb_est18:*    |
| 19:  | gb_est19:*    |
| 20:  | gb_est20:*    |
| 21:  | gb_est21:*    |
| 22:  | gb_est22:*    |
| 23:  | gb_est23:*    |
| 24:  | gb_est24:*    |
| 25:  | gb_est33:*    |
| 26:  | gb_est34:*    |
| 27:  | gb_est35:*    |
| 28:  | gb_est36:*    |
| 29:  | gb_est37:*    |
| 30:  | gb_est38:*    |
| 31:  | gb_est39:*    |
| 32:  | gb_est40:*    |
| 33:  | em_estba:*    |
| 34:  | em_estfun:*   |
| 35:  | em_esthum1:*  |
| 36:  | em_esthum2:*  |
| 37:  | em_esthum3:*  |
| 38:  | em_esthum4:*  |
| 39:  | em_esthum5:*  |
| 40:  | em_esthum6:*  |
| 41:  | em_esthum7:*  |
| 42:  | em_esthum8:*  |
| 43:  | em_esthum9:*  |
| 44:  | em_esthum10:* |
| 45:  | em_esthum11:* |
| 46:  | em_esthum12:* |
| 47:  | em_esthum13:* |
| 48:  | em_esthum14:* |
| 49:  | em_esthum15:* |
| 50:  | em_esthum16:* |
| 51:  | em_esthum17:* |
| 52:  | em_esthum18:* |
| 53:  | em_esthum19:* |
| 54:  | em_esthum20:* |
| 55:  | em_esthum21:* |
| 56:  | em_esthum22:* |
| 57:  | em_esthum23:* |
| 58:  | em_esthum24:* |
| 59:  | em_esthum25:* |
| 60:  | em_esthum26:* |
| 61:  | em_esthum27:* |
| 62:  | em_esthum28:* |
| 63:  | em_estin1:*   |
| 64:  | em_estin2:*   |
| 65:  | em_estin3:*   |
| 66:  | em_estin4:*   |
| 67:  | em_estin5:*   |
| 68:  | em_estomi:*   |
| 69:  | em_estom2:*   |
| 70:  | em_estov1:*   |
| 71:  | em_estov2:*   |
| 72:  | em_estpl1:*   |
| 73:  | em_estpl2:*   |
| 74:  | em_estpl3:*   |
| 75:  | em_estpl4:*   |
| 76:  | em_estpl5:*   |
| 77:  | em_estpl6:*   |
| 78:  | em_estpl7:*   |
| 79:  | em_estpl8:*   |
| 80:  | em_estpl9:*   |
| 81:  | em_estpl10:*  |
| 82:  | em_estro1:*   |
| 83:  | em_estro2:*   |
| 84:  | em_estro3:*   |
| 85:  | em_estro4:*   |
| 86:  | em_estro5:*   |
| 87:  | em_estro6:*   |
| 88:  | em_estro7:*   |
| 89:  | em_estro8:*   |
| 90:  | em_estro9:*   |
| 91:  | em_estro10:*  |
| 92:  | em_estro11:*  |
| 93:  | em_estro12:*  |
| 94:  | em_estro13:*  |
| 95:  | em_estro14:*  |
| 96:  | em_estro15:*  |
| 97:  | em_estro16:*  |
| 98:  | em_estro17:*  |
| 99:  | em_estro18:*  |
| 100: | em_estro19:*  |
| 101: | em_estro20:*  |
| 102: | gb_est25:*    |
| 103: | gb_est26:*    |
| 104: | gb_est27:*    |
| 105: | gb_est28:*    |
| 106: | gb_est29:*    |
| 107: | gb_est30:*    |
| 108: | gb_est31:*    |
| 109: | gb_est32:*    |
| 110: | gb_est41:*    |
| 111: | gb_est42:*    |
| 112: | gb_est43:*    |
| 113: | gb_est44:*    |
| 114: | gb_est45:*    |
| 115: | gb_est46:*    |
| 116: | gb_est47:*    |

117: gb\_est48:\*  
 118: gb\_est49:\*  
 119: gb\_est50:\*  
 120: gb\_est51:\*  
 121: gb\_est52:\*  
 122: gb\_est53:\*  
 123: gb\_est54:\*  
 124: gb\_est55:\*  
 125: gb\_est56:\*  
 126: gb\_est57:\*  
 127: gb\_est58:\*  
 128: gb\_est59:\*  
 129: gb\_est60:\*  
 130: gb\_est61:\*  
 131: gb\_est62:\*  
 132: gb\_est63:\*  
 133: gb\_est64:\*  
 134: gb\_est65:\*  
 135: gb\_est66:\*  
 136: gb\_est67:\*  
 137: gb\_est68:\*  
 138: gb\_est69:\*  
 139: gb\_est70:\*  
 140: gb\_est71:\*  
 141: gb\_est72:\*  
 142: gb\_est73:\*  
 143: gb\_est74:\*  
 144: gb\_est75:\*  
 145: gb\_est76:\*  
 146: gb\_est77:\*  
 147: gb\_est78:\*  
 148: gb\_est79:\*  
 149: gb\_est80:\*  
 150: gb\_est81:\*  
 151: gb\_est82:\*  
 152: gb\_est83:\*  
 153: gb\_est84:\*  
 154: gb\_est85:\*  
 155: gb\_est86:\*  
 156: gb\_est87:\*  
 157: gb\_est88:\*  
 158: gb\_est89:\*  
 159: gb\_est90:\*  
 160: gb\_est91:\*  
 161: gb\_est92:\*  
 162: gb\_est93:\*  
 163: gb\_est94:\*  
 164: gb\_est95:\*  
 165: gb\_est96:\*  
 166: gb\_est97:\*  
 167: gb\_est98:\*  
 168: gb\_est99:\*  
 169: gb\_est100:\*  
 170: gb\_est101:\*  
 171: gb\_est102:\*  
 172: gb\_est103:\*  
 173: gb\_est104:\*  
 174: gb\_est105:\*  
 175: gb\_est106:\*  
 176: gb\_est107:\*  
 177: gb\_est108:\*  
 178: gb\_est109:\*  
 179: gb\_est110:\*  
 180: gb\_est111:\*  
 181: gb\_est112:\*  
 182: gb\_est113:\*  
 183: gb\_est114:\*  
 184: gb\_est115:\*  
 185: gb\_est116:\*  
 186: gb\_est117:\*  
 187: gb\_est118:\*  
 188: gb\_est119:\*  
 189: gb\_est120:\*

190: gb\_est121:\*  
 191: gb\_est122:\*  
 192: gb\_est123:\*  
 193: gb\_est124:\*  
 194: gb\_est125:\*  
 195: gb\_est126:\*  
 196: gb\_est127:\*  
 197: gb\_est128:\*  
 198: gb\_est129:\*  
 199: gb\_est130:\*  
 200: gb\_est131:\*  
 201: gb\_est132:\*  
 202: gb\_est133:\*  
 203: gb\_est134:\*  
 204: gb\_est135:\*  
 205: gb\_est136:\*  
 206: gb\_est137:\*  
 207: gb\_est138:\*  
 208: gb\_est139:\*  
 209: gb\_est140:\*  
 210: gb\_est141:\*  
 211: gb\_est142:\*  
 212: gb\_est143:\*  
 213: gb\_est144:\*  
 214: gb\_est145:\*  
 215: gb\_est146:\*  
 216: gb\_est147:\*  
 217: gb\_est148:\*  
 218: gb\_est149:\*  
 219: gb\_est150:\*  
 220: gb\_est151:\*  
 221: gb\_est152:\*  
 222: gb\_est153:\*  
 223: gb\_est154:\*  
 224: gb\_est155:\*  
 225: gb\_est156:\*  
 226: gb\_est157:\*  
 227: gb\_est158:\*  
 228: gb\_est159:\*  
 229: gb\_est160:\*  
 230: gb\_est161:\*  
 231: gb\_est162:\*  
 232: gb\_est163:\*  
 233: gb\_est164:\*  
 234: gb\_est165:\*  
 235: gb\_est166:\*  
 236: gb\_est167:\*  
 237: gb\_est168:\*  
 238: gb\_est169:\*  
 239: gb\_est170:\*  
 240: gb\_est171:\*  
 241: gb\_est172:\*  
 242: gb\_est173:\*  
 243: gb\_est174:\*  
 244: gb\_est175:\*  
 245: gb\_est176:\*  
 246: gb\_est177:\*  
 247: gb\_est178:\*  
 248: gb\_est179:\*  
 249: gb\_est180:\*  
 250: gb\_est181:\*  
 251: gb\_est182:\*  
 252: gb\_est183:\*  
 253: gb\_est184:\*  
 254: gb\_est185:\*  
 255: gb\_est186:\*  
 256: gb\_est187:\*  
 257: gb\_est188:\*  
 258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| C 1        | 40    | 6.9         | 512    | 224   | AQ121628 HS_3089_A  |
| C 2        | 38.2  | 6.6         | 413    | 3     | AA171666 zo94g01.r  |
| C 3        | 38.2  | 6.6         | 516    | 8     | AA516896 vnr97e09.r |
| C 4        | 36.8  | 6.3         | 498    | 231   | AQ615572 HS_5144_B  |
| C 5        | 36.8  | 6.3         | 518    | 121   | AW818171 CM1-ST027  |
| C 6        | 35.6  | 6.1         | 318    | 224   | AQ089903 HS_3001_A  |
| C 7        | 35.4  | 6.1         | 460    | 104   | AI942816 fc67b05.x  |
| C 8        | 35.4  | 6.1         | 955    | 137   | BF603735 HVSMEH008  |
| C 9        | 35.2  | 6.0         | 345    | 243   | AQ401897            |
| C 10       | 35.2  | 6.0         | 793    | 30    | AV400883            |
| C 11       | 35.2  | 6.0         | 848    | 233   | AQ746046 HS_2277_A  |
| C 12       | 34.6  | 5.9         | 487    | 228   | AQ418525 RPCI-11-2  |
| C 13       | 34.6  | 5.9         | 555    | 228   | AQ417598 RPCI-11-2  |
| C 14       | 34.6  | 5.9         | 1101   | 219   | CNS016H0            |
| C 15       | 34    | 5.8         | 443    | 223   | AQ010299 HS_2172_B  |
| C 16       | 33.8  | 5.8         | 410    | 233   | AQ807176            |
| C 17       | 33.8  | 5.8         | 604    | 219   | AG019820            |
| C 18       | 33.8  | 5.8         | 772    | 229   | AQ488982            |
| C 19       | 33.8  | 5.8         | 894    | 221   | CNS032C0            |
| C 20       | 33.6  | 5.8         | 326    | 223   | AQ008775            |
| C 21       | 33.6  | 5.8         | 418    | 231   | AQ632834            |
| C 22       | 33.6  | 5.8         | 605    | 229   | AQ481591            |
| C 23       | 33.6  | 5.8         | 658    | 224   | AQ077737            |
| C 24       | 33.6  | 5.8         | 749    | 241   | AZ331069            |
| C 25       | 33.4  | 5.7         | 348    | 227   | AQ348508            |
| C 26       | 33.4  | 5.7         | 411    | 17    | A1226907            |
| C 27       | 33.4  | 5.7         | 517    | 230   | AQ565229 HS_5355_B  |
| C 28       | 33.4  | 5.7         | 532    | 230   | AQ565203 HS_5355_B  |
| C 29       | 33.4  | 5.7         | 637    | 227   | AQ377093            |
| C 30       | 33.4  | 5.7         | 751    | 233   | AQ745275            |
| C 31       | 33.4  | 5.7         | 764    | 226   | AQ240040            |
| C 32       | 33.4  | 5.7         | 785    | 192   | AK002866            |
| C 33       | 33.2  | 5.7         | 459    | 239   | AZ141666            |
| C 34       | 33.2  | 5.7         | 617    | 137   | BE585743            |
| C 35       | 33.2  | 5.7         | 690    | 137   | BE585745            |
| C 36       | 33.2  | 5.7         | 840    | 235   | AQ896731            |
| C 37       | 33    | 5.7         | 311    | 26    | AV223449            |
| C 38       | 33    | 5.7         | 503    | 233   | AQ800876            |
| C 39       | 33    | 5.7         | 813    | 239   | AZ190921            |
| C 40       | 32.8  | 5.6         | 224    | 159   | N49654              |
| C 41       | 32.8  | 5.6         | 471    | 225   | AQ229338            |
| C 42       | 32.8  | 5.6         | 493    | 114   | AW329763            |
| C 43       | 32.8  | 5.6         | 647    | 119   | AW695747            |
| C 44       | 32.8  | 5.6         | 649    | 151   | BF636355            |
| C 45       | 32.6  | 5.6         | 502    | 225   | AQ209803            |

## ALIGNMENTS

|            |  |            |   |             |
|------------|--|------------|---|-------------|
| RESULT 1   | AQ121628/c   | 512-bp     | DNA   | 22-SEP-1998 |
| LOCUS      | HS_3089_A1_B03_MF  | CIT        | Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3089 Col=5 Row=C, DNA sequence. |             |
| DEFINITION | AQ121628   |            |   |             |
| ACCESSION  | AQ121628.1   | GI:3498794 |   |             |
| VERSION    | GSS.   |            |   |             |
| KEYWORDS   | Homo sapiens   |            |   |             |
| SOURCE     | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.                           |            |   |             |
| REFERENCE  | 1 (bases 1 to 512)   |            |   |             |
| AUTHORS    | Mahairas, G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L. |            |   |             |
| TITLE      | Sequence-tagged connectors: A sequence approach to mapping and   |            |   |             |

## JOURNAL MEDLINE COMMENT

scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3089 row: C column: 5

Class: BAC ends

High quality sequence stop: 512.

## FEATURES

source Location/Qualifiers

1..512

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=3089 Col=5 Row=C"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 174 a 85 c 115 g 136 t 2 others

## ORIGIN

Query Match 6.9%; Score 40; DB 224; Length 512;

Best Local Similarity 59.8%; Pred. No. 0.15; Indels 0; Gaps 0;

Matches 67; Conservative 0; Mismatches 45;

QY 470 atttgggttcgaccatgtgcattgtactgtatttgcgaatgctcctgctgatttga 529

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 179 ATCTTATTGCTTCCATCTGCTTTACTCCACTTTACCAATGCTTTTAAAGTATCCCA 120

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 530 aaatcgcagcaggacagataaaatgtaattttaaagtatgtgcatt 581

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 119 AACTGAACATAGATGATCAGACTCAGGAAGATTAATGAAGACACATT 68

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

## RESULT 2

## AA171666

## LOCUS

## DEFINITION

zo94g01.r1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA

clone IMAGE:594576 5' similar to IR:G469885 G469885 EF-HAND

PROTEIN: ?, mRNA sequence.

ACCESSION AA171666

VERSION AA171666.1

KEYWORDS EST.

SOURCE human.

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

## 1 (bases 1 to 413)

## Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

## Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

## J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

## White, Y., Wyllie, T., Waterston, R. and Wilson, R.

## WashU-NCI Human EST Project

## Unpublished (1997)

## Contact: Wilson RK

## Washington University School of Medicine

## 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

## Tel: 314 286 1800

## Fax: 314 286 1810

## Email: est@wustl.edu

## This clone is available royalty-free through LLNL; contact the

## IMAGE Consortium (info@image.llnl.gov) for further information.

## Insert Length: 1500 Std Error: 0.00

## Seq primer: -28M13 rev2 from Amersham

## High quality sequence stop: 318.

## Location/Qualifiers

## 1..413

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:594576"
/clone_lib="Stratagene ovarian cancer (#937219)"
/sex="female"
/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; cloned unidirectionally. Primer: Oligo dt. Papillary
serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor
sequence: 5' CTCAGATTTTATTTTATTTT 3'"
BASE COUNT 130 a 56 c 101 g 120 t 6 others
ORIGIN
```

```
Query Match 6.6%; Score 38.2; DB 3; Length 413;
Best Local Similarity 54.78; Pred. No. 0.51;
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 75 ccaggcgatccacgaagatcggtgtacacggagctcggtggaatggacaaagatggcg 134
Db 140 CnCTGGATGTCACAGAGGAGTCCCTTTAGGAGCTTCACTTTAAAGGACAAAGCGA 199

QY 135 cgcgtgaatccgttccattccattccatatacacaaattggcagcttgatgctgcgat 194
Db 200 TTTGAAAGCCTAACACAGGATTCAGTCCCGGTTTGAGCTTGAAGAATTATTGCTTT 259

QY 195 cggagcgcagcgagcat 213
Db 260 TGAGCATCTGAGAGATT 278
```

```
RESULT 3
AA516896 516 bp mRNA EST 14-JUL-1997
LOCUS v97e09_r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
DEFINITION IMAGE:902248 5' similar to SW_RS9_HUMAN P46781 40S RIBOSOMAL
PROTEIN S9. ; mRNA sequence.
ACCESSION AA516896
VERSION AA516896.1 GI:2256355
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 516)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterson, R.
The WashU-HMMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:522312
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 473.
Location/Qualifiers
1. .516
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:902248"
```

FEATURES  
source

```
/clone_lib="Barstead mouse myotubes MPLRB5"
/cell_line="C2C12"
/lab_host="DH10B"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACGAATCTGAAGTGGAGCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATTCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified p773 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."
BASE COUNT 117 a 126 c 154 g 118 t 1 others
ORIGIN
```

```
Query Match 6.6%; Score 38.2; DB 8; Length 516;
Best Local Similarity 51.8%; Pred. No. 0.54;
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 39 attgtgaggaacctgatgaactcgcctggaagaccagcggaataccgaagatgct 98
Db 38 ATTGTGAGCANAACACTATGTGAACCCACGGGAGAACCCCTTCGGAGAAAGTTGGCGTCT 97

QY 99 tgtacacgagcttcggtggaatggacaagatggcgcgtgaaataccgttccattccc 158
Db 98 TGGACACGAGGAGCTAAAGTTGATTGGAGAGTATGACTCCGGAACAACCTGAGGTTTGA 157

QY 159 atacacaattggcagctggaatgctgtctgcgtcgagagac 202
Db 158 GGGTCAAGTTTACCTGGCCAGAGATCGTAAGCGGCCGCGGAGC 201
```

```
RESULT 4
AA615572 498 bp DNA GSS 15-JUN-1999
LOCUS HS_5144_B1_B02.SP6E RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=720 Col=3 Row=D, DNA sequence.
ACCESSION AA615572
VERSION AA615572.1 GI:5076848
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 498)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 720 row: D column: 3
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 498.
Location/Qualifiers
```

FEATURES



## source

1. .498  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-720 Col-3 Row-D"  
 /clone\_lib="RPC1-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACE3.6 vector at EcoRI sites"  
 BASE COUNT 171 a 107 c 87 g 113 t 20 others  
 ORIGIN

Query Match 6.3%; Score 36.8; DB 231; Length 498;  
 Best Local Similarity 56.7%; Pred. No. 1.5;  
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
 QY 424 gtaagaacaggcattcagtcgacaaatacgaaccccttagttatcttggttcgac 483  
 Db 271 GTTCTGGCCAGGCAATCAGGCAAGCAAGAAATAAGGGTATTAGTAGGAAGAG 330  
 QY 484 catgtgatctactgtatttgcaatgtctgctgatttgaataatcagcaggac 543  
 Db 331 GAAGTCAAAATTTGCTCTGTTCGAGATGACATGATTGTTTATTAGAAATCCCATCGAC 390

## RESULT 5

AW818171 518 bp mRNA EST 17-MAY-2000  
 LOCUS CM1-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION AW818171  
 ACCESSION AW818171  
 VERSION AW818171.1 GI:7911165  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 518)

## REFERENCE

AUTHORS  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## JOURNAL

MEDLINE 20202663  
 COMMENT  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM1-ST0277-161>)  
 299-070-h05&t3=1959-12-16&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 76

High quality sequence stop: 183.

Location/Qualifiers

## FEATURES

1. ..518

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="ST0277"

/dev\_stage="Adult"

/note="Organ: stomach; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products

## BASE COUNT

ORIGIN 193 a 119 c 89 g 117 t

## Query Match

Best Local Similarity 61.5%; Score 36.8; DB 121; Length 518;  
 Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 347 aggaacccaaacgaatttaattctcgtatgtaccagtgtaatttagatttaatt 406  
 Db 168 AGGATTAAAAACCAATTTGATCTCACATGACACAGCAAGCAAAATATAGTACACAGATT 227  
 QY 407 atgccctgttagaatgtgttaagaacaggcgatca 442  
 Db 228 GTGCTCAAGTAACGCTCTTAACACCTGCCCAATCA 263

## RESULT 6

AW818171 318 bp DNA GSS 26-AUG-1998  
 LOCUS HS\_3001\_A2\_H04\_MR CIT Approved Human Genomic Sperm Library D Homo  
 DEFINITION sapiens genomic clone Plate-3001 Col-8 Row-O, DNA sequence.  
 ACCESSION AW818171  
 VERSION AW818171.1 GI:3458814  
 KEYWORDS GSS.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 318)  
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
 Hood, L.

## REFERENCE

AUTHORS  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 99380589

## JOURNAL

MEDLINE 99380589  
 COMMENT  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3001 row: 0 column: 8

Class: BAC ends

High quality sequence stop: 318.

## FEATURES

source

1. .318

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-3001 Col-8 Row=0"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT 77 a 66 c 66 g 108 t 1 others

## Query Match

Best Local Similarity 6.1%; Score 35.6; DB 224; Length 318;  
 Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 424 gtaagaacaggcattcagtcgacaaatacgaaccccttagttatcttggttcgac 483

Db 212 GTTCTGGCCAGGCAATCAGGCAAGCAAGAAATAAGGGTATTATTTAGGAAGAG 153



QY 242 acggaggagacgggaagtcggtgttgccaaatctcgactcaagcccgactggttct 301  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 585 ACGGGGACTGGGAACCCCTGGGTACCAACAAATGCGCTTGAACACACATCCCTTTTC 644  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 302 acacatgtgtggaaccgaagacgacgtattgtaccagatcatcgagaccacaaacgaa 361  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 645 ACAGTTGGCGTATACAAAAAGGCGGACGATTCCCTTCCAAAGTGCCCATCTGAA 704  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 362 attaatcctcgatgaccaggtgtaatt 391  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 705 GGGNAATGGGAATGGAGACGCTAATATT 734  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 9  
 AZ401897/c  
 LOCUS  
 DEFINITION IM0168N22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0168N22 R, DNA sequence.

ACCESSION AZ401897  
 VERSION AZ401897.1 GI:10516971  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 345)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D.,Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0168 row: N column: 22  
 Seq primer: CACACAGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 345.  
 Location/Qualifiers

FEATURES  
 source  
 1..345  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0168N22"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42hv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."  
 BASE COUNT 104 a 77 c 58 g 106 t  
 ORIGIN

Query Match 6.0%; Score 35.2; DB 243; Length 345;  
 Best Local Similarity 55.8%; Pred. No. 4.2;  
 Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 390 tttagagatttaaatatgacctgttagaattgtttaagaacagggcattcagttcgaa 449  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 330 TTGAAGAGCTTAAATGGACGCCCTCTGAATGAGGTATATATGTTATTCGTCATGA 271  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 450 caaatacgaaccccttagttattcttggttcgacctgtgcatgtactcgtatttgcaa 509  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 270 AACCTGAGGACATCTTTTGTGTGGTTAGTTGACTCTGTACTTGCTTGCCTGCTGTTGTCAA 211  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 10  
 AV400883  
 LOCUS  
 DEFINITION AV400883 793 bp mRNA EST 05-FEB-2000  
 mori cDNA clone br--1929 T3, mRNA sequence.

ACCESSION AV400883  
 VERSION AV400883.1 GI:6904535  
 KEYWORDS EST.  
 SOURCE domestic silkworm.  
 ORGANISM Bombyx mori

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia  
 ; Bombycoidea; Bombycidae; Bombyx.

1 (bases 1 to 793)  
 Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.  
 Bombyx mori cDNA  
 Unpublished (2000)  
 Contact: Mita K  
 Genome Research Group  
 National Institute of Radiological Sciences  
 Anagawa 4-9-1, Inage, Chiba 263-8555, Japan  
 Email: kmita@nirs.go.jp

method:uni-directional, sequence direction:sequenced from T3 primer  
 (5' -> 3')  
 Project="Silkworm Genome Program in MAFF, and Research for the  
 Future Program in JSPS". see 'Silkbase',  
 <http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.  
 Location/Qualifiers

FEATURES  
 source  
 1..793  
 /organism="Bombyx mori"  
 /strain="Daizo"  
 /db\_xref="taxon:7091"  
 /clone="br--1929"  
 /clone\_lib="Bombyx mori brain Daizo P0 (just after  
 pupation)"  
 /sex="female/male mixed"  
 /tissue\_type="brain"  
 /dev\_stage="P0 (just after pupation)"  
 BASE COUNT 198 a 231 c 223 g 141 t  
 ORIGIN

Query Match 6.0%; Score 35.2; DB 30; Length 793;  
 Best Local Similarity 51.2%; Pred. No. 5.3;  
 Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 141 aaatccgttccattccatcatcacattggcagcttggaagtgtctcgcgacgcgaga 200  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 634 AAACCTGCACCCCAACCCCAAGACACCCCTCTCCAGTCCGCGCCGACGCGGTGCAA 693  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 201 gccagcgacgacgcgcacacatcctctgttatatcgagacgagacgcgggaagtc 260  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 694 CCCACAAGAGAGCTGCTTAAGGCCGACAGAAAAATAGTAAGAGATCGAGAGAGAGAGTTC 753  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

```

Qy 261 cgtgtttccaaatatctcggactcaagccgactggttc 300
      ||||| ||||| ||||| ||||| ||||| |||||
Db 754 TTAGTAATGCCAACGGTCCAGATGAAGTCCGCGGTTTC 793

RESULT 11
AQ4746046      848 bp      DNA      GSS      16-JUL-1999
LOCUS      HS_2277_A2_F11_T7C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2277 Col=22 Row=K, DNA sequence.
ACCESSION AQ4746046
VERSION AQ4746046.1 GI:5523568
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 848)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.husc.washington.edu
Plate: 2277 row: K column: 22
Seq primer: T7
Class: BAC ends
High quality sequence stop: 848.
FEATURES
    source
        1..848
        Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate=2277 Col=22 Row=K"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
            E-Coli DH10B"
BASE COUNT 228 a 136 c 117 g 144 t 223 others
ORIGIN

Query Match 6.0%; Score 35.2; DB 233; Length 848;
Best Local Similarity 55.8%; Pred. No. 5.4;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 424 gtaagaacaggcattcagtcggaacaaatcgaaccccttagttatctgttggttcgac 483
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 GTTCTGCCAGGGCAATCAGGTAGGAGAAAGAAATAACGGTATTCAATTAGGAAAAGAG 393

Qy 484 catgtgcatgtactcgtattgccaatgctcgtcctgatttgaataatcagcagggac 543
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 GAAGTCAAAATGTCCTGTTTGGAGATGACATGATTGTATATTAGAAAACCCCATGTGC 453

RESULT 12
AQ418525      487 bp      DNA      GSS      23-MAR-1999
LOCUS      RPCI-11-20318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-20318,
DEFINITION DNA sequence.
ACCESSION AQ418525
VERSION AQ418525.1 GI:4476249
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 555)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 487)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-20318.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbettig@ig.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES
    source
        1..487
        Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="GDB:7577767"
            /db_xref="taxon:9606"
            /clone="RPCI-11-20318"
            /clone_lib="RPCI-11"
            /sex="Male"
            /cell_type="lymphocytes"
            /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
            RPCI11 Human Male BAC Library"
BASE COUNT 165 a 102 c 96 g 124 t
ORIGIN

Query Match 5.9%; Score 34.6; DB 228; Length 487;
Best Local Similarity 54.3%; Pred. No. 7.2;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 410 cctgtttagaatgtttaagaacaggcattcagttcggacaatacgaaccccttagtt 469
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 CTCCTGGTAGAATTCGGTCGCGTCCAGGCAATCAGGAGGAGAAAGAAATAAGGGGTTC 186

Qy 470 atcttggttcaccatgctgcatgtactcgtattgccaatgctcctgctgattattga 529
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 AATTAGGAAAAGAGGAAGTCAAAATGTCCTGTTTGGCAGATGACATGTTGTATCTAC 246

Qy 530 aaatcagca 538
      ||||| ||
Db 247 AAAACCCCA 255

RESULT 13
AQ417598      555 bp      DNA      GSS      23-MAR-1999
LOCUS      RPCI-11-203C8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203C8,
DEFINITION DNA sequence.
ACCESSION AQ417598
VERSION AQ417598.1 GI:4475441
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 555)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

```



Thu Oct 25 13:08:12 2001

/clone="Plate-2172 Col=15 Row=N"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"  
BASE COUNT 172 a 91 c 72 g 107 t 1 others  
ORIGIN

Query Match 5.8%; Score 34; DB 223; Length 443;  
Best Local Similarity 55.7%; Pred. No. 11;  
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
Qy 424 gtaaagaacagggcattcagttcggacaaatacgaaccccttagttattcttggttcgac 483  
Db 318 GTTTTGGCCAGGGCAGTCAGGCAAGAGAGAAATAAAGCGTATTCAATAAGGAGAGAA 377  
Qy 484 catgtgcattgtactgtatttgccaattgtcctgcctgattatttgaaaaatcagca 538  
Db 378 GAAGTCAAATGCCCATGTTTGCAGATGACATGATTGTTTATNTAGAAAAACCCCA 432

Search completed: October 24, 2001, 13:13:46  
Job time: 11811 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:23 ; Search time 180.6 seconds  
(without alignments)  
611.121 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_409\_991

Perfect score: 583  
Sequence: 1 actcgtcctggagtcataag.....ttaaagtatgtcatctaa 583

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6C\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/6D\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 116.4 | 20.0        | 1096   | 2  | US-08-418-071-4   |
| 2          | 114.8 | 19.7        | 1091   | 2  | US-08-418-071-5   |
| 3          | 102.2 | 17.5        | 1106   | 2  | US-08-418-071-3   |
| 4          | 99.2  | 17.0        | 1017   | 4  | US-08-793-634B-6  |
| 5          | 91.2  | 15.6        | 1022   | 4  | US-08-793-634B-2  |
| 6          | 62.8  | 10.8        | 1110   | 1  | US-08-202-186-14  |
| 7          | 59.6  | 10.2        | 982    | 3  | US-08-973-068-28  |
| 8          | 59.6  | 10.2        | 1110   | 1  | US-08-202-186-11  |
| 9          | 59.6  | 10.2        | 1111   | 1  | US-08-202-186-9   |
| 10         | 59.6  | 10.2        | 1111   | 1  | US-08-202-186-12  |
| 11         | 59.6  | 10.2        | 1111   | 1  | US-08-202-186-24  |
| 12         | 58.6  | 10.1        | 1103   | 1  | US-08-202-186-16  |
| 13         | 58.6  | 10.1        | 1105   | 1  | US-08-202-186-18  |
| 14         | 58    | 9.9         | 1109   | 1  | US-08-202-186-13  |
| 15         | 58    | 9.9         | 1111   | 1  | US-08-202-186-15  |
| 16         | 57    | 9.8         | 1104   | 1  | US-08-202-186-17  |
| 17         | 56.4  | 9.7         | 1111   | 1  | US-08-202-186-10  |
| 18         | 46.4  | 8.0         | 287    | 2  | US-08-418-071-1   |
| 19         | 44.2  | 7.6         | 300    | 2  | US-08-418-071-2   |
| 20         | 32.2  | 5.5         | 7218   | 3  | US-08-232-463-14  |
| 21         | 31    | 5.3         | 1212   | 3  | US-09-046-578-5   |
| 22         | 30.4  | 5.2         | 1212   | 3  | US-09-046-578-1   |
| 23         | 29.4  | 5.0         | 8920   | 4  | US-08-446-855A-1  |
| 24         | 29.4  | 5.0         | 8920   | 4  | US-09-150-741-1   |
| 25         | 29.2  | 5.0         | 1209   | 6  | 5352575-4         |
| 26         | 28.8  | 4.9         | 10607  | 1  | US-08-078-090-3   |
| 27         | 28.6  | 4.9         | 1593   | 2  | US-08-524-828-2   |
|            |       |             |        |    | Sequence 2, Appli |

## ALIGNMENTS

## RESULT 1

US-08-418-071-4  
; Sequence 4, Application US/08418071  
; Patent No. 5846705

; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh

; APPLICANT: You, Li-Ru

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS

; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: United States

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 5.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/418,071

; FILING DATE: 06-APR-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley Jr., James F.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: DCB-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 596-9000

; TELEFAX: (212) 596-9090

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1096 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: circular

; MOLECULE TYPE: genomic DNA

; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Banana Bunchy Top Virus (BBTV)

US-08-418-071-4

Sequence 2, Appli  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 9, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 4, Appli





APPLICANT: You, Li-Ru  
APPLICANT: Soong, Tai-Seng  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING BANANA BUNCHY TOP VIRUS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/418,071  
FILING DATE: 06-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: DCB-1  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1106 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
MOLECULE TYPE: genomic DNA  
DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"  
HYPOPHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Banana Bunchy Top Virus (BBTV)  
US-08-418-071-3

Query Match 17.5%; Score 102.2; DB 2; Length 1106;  
Best Local Similarity 53.3%; Pred. No. 6.4e-25;  
Matches 215; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 169 tggcagcttgaagtgtctgtcgatcgagagccagcgagcgatcgcaaatctctgg 228  
DB 524 TGGCAGATTCATTACCGGAGCAATTGACGAGGAAACCGGATGATGCAAGCATCTGG 583  
QY 229 atatcgcgagacgagcgagggagtcctgtgttgccaaatctctcgactcaag 288  
DB 584 GTCTATGTCCTTATGGTAATGAGGTAAATCAACATATGGAAGTCACTAATCAAGAAG 643  
QY 289 ccgacgtgtctacacatggttggaaccagaaagacgtattgtaccagtcacag 348  
DB 644 GATTGGTCTACACCGGGTGGGAAGGAGGAGAAATATCTATTCTCTATGTGGAGAA 703  
QY 349 gaccacaaacgaatttaactcctcgatgtaccaggtgttaatttagagttattcaattat 408  
DB 704 GGATGTGACAAGCATATAGTATTGATATTCCTCGTTGTAATCAGGATTAATTAATTAAT 763  
QY 409 gccctgttagaattgttaagaacagggtcattccagttcgagacaaaccccttagt 468  
DB 764 GATGTAATAGAGGCATTAAGAGGATAGGGTTATAGAGAGTACTAAATACAAACCAATAAG 823  
QY 469 taacttggtttcgaccatgtgcagtactgtatttgccaaatgtcctgctgtatttg 528  
DB 824 ATAGTTGAATAGGTAATACATGTATGTCGTCATGCGGAATTTTCATGCTGACTTCTGT 883  
QY 529 aaaatcagcaggagacagaataaaaactgtggaattatttaaagta 571

DB 884 AAAATCTCCGAAGATCGAATAAAAAATCAATTTATTGCTGAAGAA 926

## RESULT 4

US-08-793-634B-6  
Sequence 6, Application US/08793634B  
Patent No. 6211431  
GENERAL INFORMATION:  
APPLICANT: Boevink, Petra C.  
APPLICANT: Surin, Brian P.  
APPLICANT: Keese, Paul K.  
APPLICANT: Chu, Paul W.G.  
APPLICANT: Waterhouse, Peter M.  
APPLICANT: Khan, Rafiqul I.  
APPLICANT: Larkin, Philip J.  
APPLICANT: Taylor, William C.  
APPLICANT: Marchall, Jerry S.  
TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,634B  
FILING DATE: June 9, 1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 10530  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1017 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-793-634B-6

Query Match 17.0%; Score 99.2; DB 4; Length 1017;  
Best Local Similarity 54.1%; Pred. No. 6.4e-24;  
Matches 225; Conservative 0; Mismatches 188; Indels 3; Gaps 1;

QY 163 cacaattggcagcttgaagtgtctgtcgatcgagagccagcgagcgatcgcaaatc 222  
DB 498 CCCCCTTGGCAGAGAAAGTGTATGTTAGAGGAAGAACCCAGATTATAGAAGCATA 557  
QY 223 ctctggatattcgagcagcgagcgagcgagcgagtcctgtgttgccaaatctctcgga 282  
DB 558 ATCTGGGTGTATGGACCTTGCTGTTAATGAAGCAAAATCTACATTTGCAACACATCTGCA 617  
QY 283 ctaagcccgactggttctacacatgtgttggaaccagaaagacgtattgtaccagtac 342  
DB 618 TTGAAAGATGTTGGGTATCTGCTGGAGGAAAGACACAAAGATATGATGCATCTGTG 677  
QY 343 atcagagaccccaaacgaatttaactcctcgatgtaccaggtgttaatttagatttta 402  
DB 678 ACTGCTGAGCCCTAAGAAATATGGGTATTGACATACCCAGAGTTAGTTCAGAGTATGTG 737

|     | Query Match  | 15.6%           | Score 91.2;        | DB 4;  | Length 1022; |
|-----|--|-----------------|--------------------|--------|--------------|
|     | Best Local Similarity  | 53.5%;          | Prod. No. 3.3e-21; |        |              |
|     | Matches 214;   | Conservative 0; | Mismatches 183;    | Indels | Gaps         |
| 169 | tggcagcttgaagtgtctgtcgatcgagagccgagcgagatcgacacatctctgg      | 228             |                    |        |              |
| 529 | TGGCAAAATTGAGCTTCACGAGGATTTAATGCGGGAACAGATGATCGGAGTATCATCTGG | 588             |                    |        |              |
| 229 | atatgcgagcagcagcagagacggaagtccgtgtgtgccaaatctatcggactctcaag  | 288             |                    |        |              |
| 598 | CTCATGTGTTACACGAGAGAGGAAACACAGCTTCGCGAAGGAAT---AATCAGG       | 645             |                    |        |              |

[illegible]





Db 790 AGGATTATTAAATGATGGTTATTAGAGAAATTAAGAAATCGAATTAATTCAAAGCGCGA 849  
QY 452 aatacgaacccctagtattcttgggttcgaccatgctgactgctgactgctgactgctgactg 511  
Db 850 AATATGAACCCGTTTGAAGATAG---TAGAATAATGCGAAGTCAATGTAATGCTAACT 906  
QY 512 tctgctgattatttgaataatcagcagcgagacagaataaaactg 555  
Db 907 TCCTTCGAAGAAGGAATCTTCTTGAAGATCGAATAAAGTTG 950

RESULT 11  
US-08-202-186-24  
; Sequence 24, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JEFFERY, Donald D.  
; REGISTRATION NUMBER: 19,980  
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202 672 5300  
; TELEFAX: 202 672 5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1111 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 129..989  
US-08-202-186-24

Query Match 10.2%; Score 59.6; DB 1; Length 1111;  
Best Local Similarity 51.3%; Pred. No. 1.7e-10;  
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 180 agtgcgtctgctgagacgagacgagacgagacgagacgagacgagacgagacgagacgagacg 239  
Db 611 AGAGTGGAGATATATCATGGCGGACCCATGTCATCGGAGAAATATTTGGTCTATGGCC 670  
QY 240 agacgagagagacgagagagacgagacgagacgagacgagacgagacgagacgagacgagacg 299  
Db 671 AAATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 730  
QY 300 ctacacatgtgtggaacccagaaagacgtattgtaccagctacatcgagaccccaaacg 359  
Db 731 TTATTCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 787

QY 360 aaatttaactcctcgatgtaaccaggtgtaatttagatatttaattatccctgtaga 419  
Db 788 ---TGTATATTGATATTCCAAGATGCAAGAGGATATTTAAATATATGGGTATTAGA 844  
QY 420 atgtgttaagaacagggcattcagttcggacaaatacgaacccctctagttatcttgggtt 479  
Db 845 GGAATTTAAGAAATGAATAATTCAAAGCGGAAATATGAACCCCTTTTGAAGATAG---T 901  
QY 480 cgaccatgtgcatgactcgtatttgcgaatgctcctgctgctgatttggaaatcagcag 539  
Db 902 AGAATATGCTCAAGTCATTTGTAATGGCTAACTTCCTTCCGAGGAGGAATCTTTTCTGA 961  
QY 540 ggcagagaataaaactg 555  
Db 962 AGATCGAATAAAGTTG 977

RESULT 12  
US-08-202-186-16  
; Sequence 16, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JEFFERY, Donald D.  
; REGISTRATION NUMBER: 19,980  
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202 672 5300  
; TELEFAX: 202 672 5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1103 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
US-08-202-186-16

Query Match 10.1%; Score 58.6; DB 1; Length 1103;  
Best Local Similarity 52.5%; Pred. No. 3.7e-10;  
Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;

QY 212 atcgacatactcttgatgacgagacgagacgagacgagacgagacgagacgagacgagacg 271  
Db 616 ATCGAAGGATATTATTGGTCTTACGCCCAATGGAGCGAAGGAAACACACGTTTGGCAA 675  
QY 272 aatatctcgagactcaagcccgactggtctacacatgtgtggaacccagaaagacgtat 331  
Db 676 AACATTTAATGAAGACTAAGAAATCGGTTTATTTCGCCAGGAGGAGGAGGAGGAGGAGGAGG 735

|     |     |   |     |
|-----|-----|---|-----|
| 332 | Qy  | tgtaccaggtacatcggaggaccacaaacgaaatttaattccctcgatgtaccaccaggtgaatt   | 391 |
|     |     |   |     |
|     |     |   |     |
| 736 | Ddb | GTAGATTGTATATTATGAGGATATA-----GTTATATTTGATATGCCAGATGCAAG            | 789 |
|     |     |   |     |
|     |     |   |     |
| 392 | Qy  | tagagttttaattaaattatgcccctggtcagaaatgtgtagaacaacaggcgattcagttcggaca | 451 |
|     |     |   |     |
|     |     |   |     |
| 790 | Ddb | AGGAATATTTAAACTATGCTCTATTAGAAAGAAATTTAAAAATGGAATTTTCAAGCGGGA        | 849 |
|     |     |   |     |
|     |     |   |     |
| 452 | Qy  | aatcacgaaccccttagttatctctgggttcgaccatgtgcattcactcglatcttgccaatg     | 511 |
|     |     |   |     |
|     |     |   |     |
| 850 | Ddb | AATTATGACCCGTTTGAANAATTG---TAGAATATGTGGAAGTCATTGTAATGGCTAACT        | 906 |
|     |     |   |     |
|     |     |   |     |
| 512 | Qy  | tcctgcctgattatttgaataatcagcaggacagaaataaaact                        | 554 |
|     |     |   |     |
|     |     |   |     |
| 907 | Ddb | TCCTTCGGAAGGAGGAATCTTTTCTGAAGATCGAATTAAGCT                          | 949 |
|     |     |   |     |
|     |     |   |     |

```

RESULT 13
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAM, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-18

```

|     | Query Match  | 10.1%        | Score 58.6;        | DB 1;     | Length 1105; |
|-----|--|--------------|--------------------|-----------|--------------|
|     | Best Local Similarity  | 52.5%;       | Prod. No. 3.7e-10; |           |              |
|     | Matches 180;   | Conservative | 0; Mismatches 154; | Indels 9; | Gaps 2;      |
| 212 | atccgacaatctctgatatcgagcagacgagcagcagcggagctcgtgtttgcc       | 271          |                    |           |              |
|     |  |              |                    |           |              |
| 618 | ACCGAAGGATAATTGGGCTATGTGCCCAATCGAGGAGGAAGACAACTTATGCA        | 677          |                    |           |              |
|     |  |              |                    |           |              |
| 272 | aatctctcggactcaagcccgactggtctcacacatggtgtgacacagaagagcgat    | 331          |                    |           |              |
|     |  |              |                    |           |              |
| 678 | AATATTAAATGAAACGAGAAATTCGCTTTTATTCGCCAGGAGGAAATCATTGGATAT    | 737          |                    |           |              |
|     |  |              |                    |           |              |
| 332 | tgtaccagtcacatcaggaccacaaacgaattaatctctcgtatgtaccacaggttaatt | 391          |                    |           |              |
|     |  |              |                    |           |              |

|     |   |               |     |
|-----|---|---------------|-----|
| 738 | GTAGATTGTAATAATTATGAGGAATA-----GTTATATTGATATTC      | CCAGATGCAAG   | 791 |
| 392 | tagagttattaaattagccctggttagaattgtttaagaaacagggcattc | agttcggaca    | 451 |
| 792 | AGGAATATTTAAACTATGCTTTTATAGAAGAATTTAGAATGGAATTA     | TTCAAAGCGGA   | 851 |
| 452 | aatcagaaccccttagttattcttgggttcgaccatgctgctactcgatt  | tttgcgaatg    | 511 |
| 852 | AATATGAAACCGTTTTGAAATTTG---TAGAATATGTGGAAGTCATT     | GTATGAGCTAACT | 908 |
| 512 | tctcgctgattatttgaataatcagcagggacagaataaaact         |               | 554 |
| 908 | TCCCTCCAGGAGGAATCTTTTCAGAAAGATCGAAATAAGCT           |               | 951 |

```

RESULT 14
US-08-202-186-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; PS-08-202-186-13

```

|   | Query Match           | 9.9%  | Score 58;        | DB 1;           | Length 1109; |      |
|---|-----------------------|---|------------------|-----------------|--------------|------|
|   | Best Local Similarity | 51.1%   | Pred. No. 6e-10; |                 |              |      |
|   | Matches 192;          | Conservative  | 0;               | Mismatches 175; | Indels       | Gaps |
| y | 180                   | agtgtgtctgcgatactcgagagccagcgagatcgacacatctcttgatattcggcacg     | 239              |                 |              |      |
| b | 584                   | AGAGGTGGAGAAATATCATGTGGCCACCCATGTCATCGGAGATAATTTGGGTCTATGGCCC   | 643              |                 |              |      |
| y | 240                   | agacggaggagacgggaaagtccgtgtttggccaaatatctcgactcaagcccgactggtt   | 299              |                 |              |      |
| b | 644                   | AAATGGAGGAGAGGAAGACAACTATGTCAAAACGCTCTAATGAAGACGAGAAATGCGTT     | 703              |                 |              |      |
| y | 300                   | ctacacatgtgttggaaccagaaaaggacgtattgtaccagtcacatcggaggaccacaaacg | 359              |                 |              |      |



us-09-462-955-1\_copy\_409\_991.rni

Thu Oct 25 13:08:11 2001

...



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:22 ; Search time 393.52 Seconds  
(without alignments)  
607.925 Million cell updates/sec

Title: US-09-462-955-1\_copy\_611\_991  
Perfect score: 381  
Sequence: 1 cagcgagcagtcgacacatc.....ttaaagtgtgtcatctaa 381

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

|     |   |
|-----|---|
| 1:  | /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.* |
| 2:  | /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.* |
| 3:  | /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.* |
| 4:  | /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.* |
| 5:  | /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.* |
| 6:  | /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.* |
| 7:  | /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.* |
| 8:  | /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.* |
| 9:  | /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.* |
| 10: | /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.* |
| 11: | /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.* |
| 12: | /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.* |
| 13: | /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.* |
| 14: | /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.* |
| 15: | /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.* |
| 16: | /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.* |
| 17: | /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.* |
| 18: | /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.* |
| 19: | /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.* |
| 20: | /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.* |
| 21: | /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.* |
| 22: | /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 112.4 | 29.5        | 1096   | AAV71833 | BTV DNA II clone   |
| 2          | 112.4 | 29.5        | 1096   | AAA38950 | Banana bunchy top  |
| 3          | 110.8 | 29.1        | 1091   | AAV71834 | BTV DNA II clone   |
| 4          | 110.8 | 29.1        | 1091   | AAA38947 | Banana bunchy top  |
| 5          | 110.6 | 29.0        | 886    | AAA34687 | DNA sequence of BB |
| 6          | 89.6  | 23.5        | 1106   | AAV71832 | BTV DNA I clone (  |
| 7          | 88.6  | 23.3        | 1017   | AAT13165 | SCSV segment 6. S  |
| 8          | 88.2  | 23.1        | 1022   | AAT13161 | SCSV segment 2. S  |
| 9          | 88    | 23.1        | 1106   | AAA38946 | Banana bunchy top  |
| 10         | 80.4  | 21.1        | 593    | AAA34686 | DNA sequence of BB |
| 11         | 62.8  | 16.5        | 1110   | AAV24089 | Banana bunchy top  |

|    |      |      |       |    |          |                    |
|----|------|------|-------|----|----------|--------------------|
| 12 | 59.6 | 15.6 | 982   | 18 | AAT49405 | Banana bunchy top  |
| 13 | 59.6 | 15.6 | 1110  | 19 | AAV24086 | Banana bunchy top  |
| 14 | 59.6 | 15.6 | 1111  | 19 | AAV24077 | Banana bunchy top  |
| 15 | 59.6 | 15.6 | 1111  | 19 | AAV24084 | Banana bunchy top  |
| 16 | 59.6 | 15.6 | 1111  | 19 | AAV24087 | Banana bunchy top  |
| 17 | 58.6 | 15.4 | 1103  | 19 | AAV24091 | Banana bunchy top  |
| 18 | 58.6 | 15.4 | 1105  | 19 | AAV24093 | Banana bunchy top  |
| 19 | 58   | 15.2 | 1109  | 19 | AAV24088 | Banana bunchy top  |
| 20 | 58   | 15.2 | 1111  | 19 | AAV24090 | Banana bunchy top  |
| 21 | 57   | 15.0 | 1104  | 19 | AAV24092 | Banana bunchy top  |
| 22 | 56.4 | 14.8 | 1111  | 19 | AAV24085 | Banana bunchy top  |
| 23 | 45.8 | 12.0 | 936   | 22 | AAF58252 | Oligonucleotide D1 |
| 24 | 45.8 | 12.0 | 936   | 22 | AAF58254 | Oligonucleotide D1 |
| 25 | 45.8 | 12.0 | 936   | 22 | AAF58257 | Oligonucleotide D1 |
| 26 | 45.8 | 12.0 | 936   | 22 | AAF58259 | Oligonucleotide D2 |
| 27 | 45.8 | 12.0 | 936   | 22 | AAF58262 | Oligonucleotide D2 |
| 28 | 45.8 | 12.0 | 938   | 22 | AAF58255 | Oligonucleotide D1 |
| 29 | 42.4 | 11.1 | 936   | 22 | AAF58252 | Oligonucleotide D1 |
| 30 | 42.4 | 11.1 | 936   | 22 | AAF58254 | Oligonucleotide D1 |
| 31 | 42.4 | 11.1 | 936   | 22 | AAF58257 | Oligonucleotide D1 |
| 32 | 42.4 | 11.1 | 936   | 22 | AAF58259 | Oligonucleotide D2 |
| 33 | 42.4 | 11.1 | 938   | 22 | AAF58262 | Oligonucleotide D2 |
| 34 | 42.4 | 11.1 | 938   | 22 | AAF58255 | Oligonucleotide D1 |
| 35 | 40.2 | 10.6 | 300   | 20 | AAV71831 | Subgenomic fragmen |
| 36 | 40.2 | 10.6 | 300   | 21 | AAA38949 | Banana bunchy top  |
| 37 | 37   | 9.7  | 244   | 22 | AAF58238 | Oligonucleotide D1 |
| 38 | 34.2 | 9.0  | 244   | 22 | AAF58238 | Oligonucleotide D1 |
| 39 | 33.8 | 8.9  | 287   | 20 | AAV71830 | Subgenomic fragmen |
| 40 | 33.8 | 8.9  | 287   | 21 | AAA38948 | Banana bunchy top  |
| 41 | 32.6 | 8.6  | 5053  | 19 | AAV22816 | Nucleotide sequenc |
| 42 | 32.6 | 8.6  | 5055  | 19 | AAV17236 | DNA from a region  |
| 43 | 31.4 | 8.2  | 41599 | 21 | AAZ35351 | Cosmid including s |
| 44 | 30.8 | 8.1  | 41599 | 21 | AAZ35351 | Cosmid including s |
| 45 | 30.6 | 8.0  | 9345  | 21 | AAZ93966 | Mouse.uromodulin p |

## ALIGNMENTS

RESULT 1

AAV71833

ID AAV71833 standard; DNA; 1096 BP.

XX AC AAV71833;

XX DT 10-FEB-1999 (first entry)

XX DE BTV DNA II clone (2-17) nucleotide sequence.

XX KW Banana bunchy top virus; BTV; PCR technique; plant; tissue; banana; Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.

XX OS Banana bunchy top virus.

XX FH Key Location/Qualifiers

FT TATA\_signal 1..7

FT stem\_loop /\*tag= a

FT CDS /\*tag= b

FT /\*tag= c

FT /\*product= "ORF-V2 product"

FT polyA\_signal /transl\_except= (pos:215..217, aa:Gly)

FT polyA\_signal 533..538

FT polyA\_signal /\*tag= d

FT polyA\_signal 799..804

FT polyA\_signal /\*tag= e

FT polyA\_signal 907..912

FT polyA\_signal /\*tag= f

FT polyA\_signal 1030..1035

FT /\*tag= ge

XX US5846705-A.



XX AAV71834;  
 XX AC 10-FEB-1999 (first entry)  
 XX DT  
 XX DE BBTV DNA II clone (2) nucleotide sequence.  
 XX KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
 XX KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
 XX OS Banana bunchy top virus.  
 XX FN US5846705-A.  
 XX PD 08-DEC-1998.  
 XX PF 06-APR-1995; 95US-0418071.  
 XX PR 06-APR-1995; 95US-0418071.  
 XX PA (BTOT-) DEV CENT BIOTECHNOLOGY.  
 XX PI Soong T, Wu R, You L;  
 XX DR WPI; 1999-059037/05.  
 XX AC Nucleic acids having banana bunchy top virus component sequences -  
 XX PT used to design primers for use in polymerase chain reaction  
 XX PT detection of the virus  
 XX PS Disclosure; Fig 12A-C; 27pp; English.  
 XX CC This represents the nucleotide sequence of a banana bunchy top virus  
 CC (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention  
 CC provides nucleic acid sequences associated with BBTV that can be used in  
 CC a PCR technique for detecting BBTV. The nucleic acid sequences (AAV71830  
 CC to AAV71833) are used as the basis for the construction of PCR primers,  
 CC to detect BBTV infection. The PCR technique is used for detecting BBTV  
 CC in plant tissues (preferably banana, especially Musa species). The  
 CC virus, one of the most important banana species viruses, causes phloem  
 CC damage and is transmitted by aphids. PCR detection gives accurate,  
 CC reliable and specific determination of absence or presence of the  
 CC virus.  
 XX SQ Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other;

Query Match 29.1%; Score 110.8; DB 20; Length 1091;  
 Best Local Similarity 57.7%; Pred. No. 5.3e-27;  
 Matches 218; Conservative 0; Mismatches 157; Indels 3; Gaps 1;  
 QY 4 cggacgacgcacaaatcctctggtatgcgagcagagagcggagagcggagtcgtgt 63  
 DB 559 ctgatgcgaagtatcatctgggtatgcgctcgcggagcgaaggaagtcacact 618  
 QY 64 ttgccaaatctcggactcaagcccgactgttctacacatgtgtgtggaaccagaaagg 123  
 DB 619 tcgcaagatatctatcaataaactggatgggtatcaacacgtgtggaagcgtcgg 678  
 QY 124 acgtattgtaccagatcacatcggaagcccaacgaataatttaactcgtatgccaggt 183  
 DB 679 atatgatgcacataacagatggatccttgataatcattgtgattgtatccccagaa 738  
 QY 184 gtaatttagagtatttaaatatgcctgttagaagtgttaagaacagggcattcagtt 243  
 DB 739 gtcattcagattatctgaattatgcgttatagaacaataaagaatagattttataa 798  
 QY 244 cggacaaatacgaacc---ccttagttatcttggttcgaccatgtgcattcgtat 300  
 DB 799 atacaaatacgaaccatgtgtgattagaaagatggacaaatgtccatgtaattgta 858  
 QY 301 ttgccaaatgctcgtcgtattttgaaatcagcaggagacagataaaactgtggaata 360  
 DB 799 atacaaatacgaaccatgtgtgattagaaagatggacaaatgtccatgtaattgta 858

DB 859 tggcaaatgtgtgctgattattgtaaaatttcagaagataaataataatt 918  
 QY 361 tttaaagtagtgtgtcatc 378  
 DB 919 gttgagaagaagaaacttc 936  
 RESULT 4  
 AAA38947  
 ID AAA38947 standard; DNA; 1091 BP.  
 XX AAA38947;  
 XX DT 23-AUG-2000 (first entry)  
 XX DE Banana bunchy top virus related nucleotide sequence #2.  
 XX KW Banana bunchy top virus; BBTV; detection; ds.  
 XX OS Banana bunchy top virus.  
 XX PN TW360710-A.  
 XX PD 11-JUN-1999.  
 XX PF 30-JUN-1994; 94TW-0106105.  
 XX PR 30-JUN-1994; 94TW-0106105.  
 XX PA (BTOT-) DEV CENT BIOTECHNOLOGY.  
 XX PI Wu R, You L, Song T;  
 XX DR WPI; 2000-316145/27.

Two circular single-stranded DNAs associated with banana bunchy top virus and detection of the virus -  
 PS Claim 2; Page 1; 7pp; Chinese.  
 XX The present invention describes two circular single-stranded DNAs  
 CC associated with banana bunchy top virus (BBTV) and the encoded proteins.  
 CC Also described is a method of detecting the BBTV virus using the  
 CC polymerase chain reaction (PCR). The present sequence represents a  
 CC BBTV related nucleotide sequence from the present invention.  
 XX SQ Sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;

Query Match 29.1%; Score 110.8; DB 21; Length 1091;  
 Best Local Similarity 57.7%; Pred. No. 5.3e-27;  
 Matches 218; Conservative 0; Mismatches 157; Indels 3; Gaps 1;  
 QY 4 cggacgacgcacaaatcctctggtatgcgagcagagagcggagagcggagtcgtgt 63  
 DB 559 ctgatgcgaagtatcatctgggtatgcgctcgcggagcgaaggaagtcacact 618  
 QY 64 ttgccaaatctcggactcaagcccgactgttctacacatgtgtgtggaaccagaaagg 123  
 DB 619 tcgcaagatatctatcaataaactggatgggtatcaacacgtgtggaagcgtcgg 678  
 QY 124 acgtattgtaccagatcacatcggaagcccaacgaataatttaactcgtatgccaggt 183  
 DB 679 atatgatgcacataacagatggatccttgataatcattgtgattgtatccccagaa 738  
 QY 184 gtaatttagagtatttaaatatgcctgttagaagtgttaagaacagggcattcagtt 243  
 DB 739 gtcattcagattatctgaattatgcgttatagaacaataaagaatagattttataa 798  
 QY 244 cggacaaatacgaacc---ccttagttatcttggttcgaccatgtgcattcgtat 300  
 DB 799 atacaaatacgaaccatgtgtgattagaaagatggacaaatgtccatgtaattgta 858

|    | Query Match               | 29.0%;   | Score 110.6;       | DB 20;    | Length 886; |
|----|---------------------------|--|--------------------|-----------|-------------|
|    | Best Local Similarity     | 58.4%;   | Pred. No. 5.6e-27; |           |             |
|    | Matches 213; Conservative | 0;   | Mismatches 149;    | Indels 3; | Gaps 1;     |
| yy | 4                         | cgaacatcgcaaatccctcctgatatcgagcagacggaggagacgcgggaagtccgtgt      | 63                 |           |             |
| bb | 285                       | ctgagtcagaagtatactcttggttacgttcccacggagcgaaagaaagtaaacct         | 344                |           |             |
| yy | 64                        | tgcacaattctcgactcaaacgccgactggtttcacacatggttgtgtgaaaccagaaggg    | 123                |           |             |
| bb | 345                       | tcgcaagatatcatataaaacctggatgggatatactacacgtgtggaagacgtcgg        | 404                |           |             |
| yy | 124                       | acgtattgtaccagtagacatcagagaccaccaaaaagaatttaactcctgatgataccacggt | 183                |           |             |
| bb | 405                       | atatgtgcacatcataaogatggatccctgataatacatctggattatctgataccccagaa   | 464                |           |             |
| yv | 184                       | qtaatttagactatttaaattatgcacctgtgaagaatggtttaagAACAGGgattcagtt    | 243                |           |             |

|       |                                  |   |
|-------|----------------------------------|---|
| b     | 465                              | gtcattcagattatctgaattatgcggttatagaacaaattaagaataagagttttaataa         |
| y     | 244                              | cggacaataacgaacc---ccctaggttacttcttggttcgcaccaatgctgatctactcgatat     |
| b     | 525                              | atacaaaaacgaaccattgctgattagaagaagatggacaaaaatgccatgtaattgttta         |
| y     | 301                              | tgcctaagtctcgtcgtgattatttgaaaaatcacgacgggacagataaaacctgtgggaata       |
| b     | 585                              | tggcaaatgctgctgattattgtaaaatttcagaagacagataaaaaataattaatt             |
| y     | 361                              | tttaa 365   |
| b     | 645                              | gttga 649   |
| <hr/> |                                  |   |
|       | RESULT                           | 6   |
|       | AAV71832                         |   |
| ID    | AAV71832 standard; DNA; 1106 BP. |   |
| XX    | AC                               | AAV71832;   |
| XX    | AC                               |   |
| XX    | DE                               | DE  |
| XX    | DE                               | BTVV DNA I clone (7-4-2) nucleotide sequence.                         |
| XX    | DE                               |   |
| KW    | KW                               | Banana bunchy top virus; BBTv; PCR technique; plant; tissue; banana;  |
| KW    | KW                               | Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; |
| KW    | KW                               |   |
| OS    | OS                               | Banana bunchy top virus.  |
| XX    | Key                              | Location/Qualifiers   |
| FT    | TATA_signal                      | 1..8  |
| FT    |                                  | /*tag= a  |
| FT    | stem_loop                        | 9...38  |
| FT    |                                  | /*tag= b  |
| FT    | CDS                              | 62..922   |
| FT    |                                  | /*tag= c  |
| FT    |                                  | /product= "ORF-V2 product"  |
| FT    |                                  | /transl_except= (pos:335..337, aa:Gly)                                |
| FT    |                                  | /transl_except= (pos:518..520, aa:Ser)                                |
| FT    | polyA_signal                     | 380..385  |
| FT    |                                  | /*tag= d  |
| FT    | polyA_signal                     | 901..906  |
| FT    |                                  | /*tag= e  |
| XX    | US5846705-A.                     |   |
| XX    | PN                               |   |
| XX    | PD                               | 08-DEC-1998.  |
| XX    | PF                               | 06-APR-1995; 95US-0418071.  |
| XX    | PP                               |   |
| XX    | PR                               | 06-APR-1995; 95US-0418071.  |
| XX    | PA                               | (BIOT-) DEV CENT BIOTECHNOLOGY.                                       |
| PI    | PI                               | Soong T, Wu R, You L;   |
| PI    | PI                               |   |
| DR    | DR                               | WP1; 1999-059037/05.  |
| DR    | DR                               | P-PSDB; AAW87459.   |
| XX    | XX                               |   |
| PT    | PT                               | Nucleic acids having banana bunchy top virus component sequences -    |
| PT    | PT                               | used to design primers for use in polymerase chain reaction           |
| PT    | PT                               | detection of the virus  |
| XX    | XX                               |   |
| PS    | PS                               | Claim 1; Fig 11A-B; 27pp; English.                                    |
| XX    | XX                               |   |
| CC    | CC                               | This represents the nucleotide sequence of a banana bunchy top virus  |
| CC    | CC                               | (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The    |
| CC    | CC                               | invention provides nucleic acid sequences associated with BBTV that   |
| CC    | CC                               | be used in a PCR technique for detecting BBTV. The nucleic acid       |
| CC    | CC                               | sequences (AAV71830 to AAV71833) are used as the basis for the        |
| CC    | CC                               | construction of PCR primers, to detect BBTV infection. The PCR tech   |













Thu Oct 25 13:08:21 2001

Oy 310 tcctgcctgattattgaaatcagcaggagacagaataaaactg 353  
||||| | | | | | | | | | | | | | | | | |  
Db 907 tccttcgaaggaaggaatcctttctgaagatcgataaagtig 950

Search completed: October 24, 2001, 10:07:24  
Job time: 629 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:46 ; Search time 5479.82 Seconds  
(without alignments)  
657.236 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_611\_991  
Perfect score: 381  
Sequence: 1 cagcgagcagtcgcacaatc.....ttaagtgatgtcatctaa 381

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*  
48: gb\_est48.\*  
49: gb\_est49.\*  
50: gb\_est50.\*  
51: gb\_est51.\*  
52: gb\_est52.\*  
53: gb\_est53.\*  
54: gb\_est54.\*  
55: gb\_est55.\*  
56: gb\_est56.\*  
57: gb\_est57.\*  
58: gb\_est58.\*  
59: gb\_est59.\*  
60: gb\_est60.\*  
61: gb\_est61.\*  
62: gb\_est62.\*  
63: gb\_est63.\*  
64: gb\_est64.\*  
65: gb\_est65.\*  
66: gb\_est66.\*  
67: gb\_est67.\*  
68: gb\_est68.\*  
69: gb\_est69.\*  
70: gb\_est70.\*  
71: gb\_est71.\*  
72: gb\_est72.\*  
73: gb\_est73.\*  
74: gb\_est74.\*  
75: gb\_est75.\*  
76: gb\_est76.\*  
77: gb\_est77.\*  
78: gb\_est78.\*  
79: gb\_est79.\*  
80: gb\_est80.\*  
81: gb\_est81.\*  
82: gb\_est82.\*  
83: gb\_est83.\*  
84: gb\_est84.\*  
85: gb\_est85.\*  
86: gb\_est86.\*  
87: gb\_est87.\*  
88: gb\_est88.\*  
89: gb\_est89.\*  
90: gb\_est90.\*  
91: gb\_est91.\*  
92: gb\_est92.\*  
93: gb\_est93.\*  
94: gb\_est94.\*  
95: gb\_est95.\*  
96: gb\_est96.\*  
97: gb\_est97.\*  
98: gb\_est98.\*  
99: gb\_est99.\*  
100: gb\_est100.\*  
101: gb\_est101.\*  
102: gb\_est102.\*  
103: gb\_est103.\*  
104: gb\_est104.\*  
105: gb\_est105.\*  
106: gb\_est106.\*  
107: gb\_est107.\*  
108: gb\_est108.\*  
109: gb\_est109.\*  
110: gb\_est110.\*  
111: gb\_est111.\*  
112: gb\_est112.\*  
113: gb\_est113.\*  
114: gb\_est114.\*  
115: gb\_est115.\*  
116: gb\_est116.\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



|  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|
| <p>Plate: 720 row: D column: 3</p> <p>Seq primer: SP6</p> <p>Class: BAC ends</p> <p>High quality sequence stop: 498.</p> <p>Location/Qualifiers</p> <p>1. .498</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone_lib="plate=720 Col=3 Row=D"</p> <p>/clone_lib="RPCI-11 Human Male BAC Library"</p> <p>/sex="male"</p> <p>/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"</p>  |  |  |  |  |  |  |  |  |  |
| <p>BASE COUNT</p> <p>171 a 107 c 87 g 113 t 20 others</p> <p>ORIGIN</p>  |  |  |  |  |  |  |  |  |  |
| <p>Query Match 9.7%; Score 36.8; DB 231; Length 498;</p> <p>Best Local Similarity 56.7%; Pred. No. 1.3;</p> <p>Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;</p>  |  |  |  |  |  |  |  |  |  |
| <p>QY 222 gtaagaacagggcattcagtcgtgcgacaaatacgaaccccttagttatcttggttcgac 281</p> <p>   </p>  |  |  |  |  |  |  |  |  |  |
| <p>Db 271 GTCTGGCCGCGCAATCAGCAACACAGAAAGAAATAAGGGGTATTAAGTTAGGAAAAGAG 330</p> <p>   </p>   |  |  |  |  |  |  |  |  |  |
| <p>QY 282 catgtgcgtactcgtattgccaatctcgtcgtattgaaatacagcaggac 341</p> <p>   </p>  |  |  |  |  |  |  |  |  |  |
| <p>Db 331 GAAGTCAAATGTCCTGTTGCAGATGACATGATGTTATTATTAGAAATCCCATCGAC 390</p> <p>   </p>  |  |  |  |  |  |  |  |  |  |
| <p>RESULT 3</p> <p>AW818171 518 bp mRNA EST 17-MAY-2000</p> <p>LOCUS CM1-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.</p> <p>DEFINITION AW818171 GI:7911165</p> <p>ACCESSION AW818171.1</p> <p>VERSION EST.</p> <p>KEYWORDS human.</p> <p>SOURCE Homo sapiens</p> <p>ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>REFERENCE 1 (bases 1 to 518)</p> <p>AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalhal,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.J.</p> <p>TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags</p> <p>JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)</p> <p>MEDLINE 20202663</p> <p>COMMENT Contact: Simpson A.J.G.</p> <p>Laboratory of Cancer Genetics</p> <p>Ludwig Institute for Cancer Research</p> <p>Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil</p> <p>Tel: +55-11-2704922</p> <p>Fax: +55-11-2707001</p> <p>Email: asimpson@ludwig.org.br</p> <p>This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (<a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-cm1-st0277-161299-070-h05&amp;st3=1999-12-16&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-cm1-st0277-161299-070-h05&amp;st3=1999-12-16&amp;t4=1</a>)</p> <p>Seq primer: puc 18 forward</p> <p>High quality sequence start: 76</p> <p>High quality sequence stop: 183.</p> <p>Location/Qualifiers</p> <p>1. .518</p> <p>/organism="Homo sapiens"</p> |  |  |  |  |  |  |  |  |  |
| <p>FEATURES</p> <p>source</p>  |  |  |  |  |  |  |  |  |  |
| <p>/db_xref="taxon:9606"</p> <p>/clone_lib="Adult"</p> <p>/note="Organ: stomach; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from OPESPEC PCR (U.S. Letters Patent application NO. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."</p>   |  |  |  |  |  |  |  |  |  |
| <p>BASE COUNT</p> <p>193 a 119 c 89 g 117 t</p> <p>ORIGIN</p>  |  |  |  |  |  |  |  |  |  |
| <p>QY 145 agggcccaaacgaaatttaattcctcgtatccacccagtgtaatttagatattaaatt 204</p> <p>   </p>  |  |  |  |  |  |  |  |  |  |
| <p>Db 168 AGGATTAAAAACCAATTTGATCTCACATGCACACAGCAAAAATATAGTAAACAAGATT 227</p> <p>   </p>  |  |  |  |  |  |  |  |  |  |
| <p>QY 205 atgcctgttagaatgtgttaagaacagggcattca 240</p> <p>   </p>   |  |  |  |  |  |  |  |  |  |
| <p>Db 228 GTGCTCAGAAGTAAGCTCTTAACACCTGCCCAATCA 263</p> <p>   </p>  |  |  |  |  |  |  |  |  |  |
| <p>RESULT 4</p> <p>AQ089903 318 bp DNA GSS 26-AUG-1998</p> <p>LOCUS HS_3001_A2_H04_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.</p> <p>DEFINITION AQ089903</p> <p>ACCESSION AQ089903.1</p> <p>VERSION GI:3459814</p> <p>KEYWORDS GSS.</p> <p>SOURCE human.</p> <p>ORGANISM Homo sapiens</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>REFERENCE 1 (bases 1 to 318)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L</p> <p>High Throughput Sequencing Center</p> <p>University of Washington</p> <p>401 Queen Anne Avenue North, Seattle, WA 98109, USA</p> <p>Tel: (206) 616-3618</p> <p>Fax: (206) 616-3887</p> <p>Email: jwallace@u.washington.edu</p> <p>Sequence Tagged Connector</p> <p>Plate: 3001 row: O column: 8</p> <p>Class: BAC ends</p> <p>High quality sequence stop: 318.</p> <p>Location/Qualifiers</p> <p>1. .318</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone_lib="Plate=3001 Col=8 Row=O"</p> <p>/clone_lib="CIT Approved Human Genomic Sperm Library D"</p> <p>/sex="male"</p> <p>/note="organ: sperm; Vector: pBelobAC11;</p>   |  |  |  |  |  |  |  |  |  |

|    | Matches | 71: Conservative   | 0: Mismatches | 59: Indels | 0: Gaps |
|----|---------|--|---------------|------------|---------|
| QY | 222     | gttaagaacaggcattcagcttoggacaataacgaacccttagttatcttgggttcac | 281           |            |         |
|    |         |  |               |            |         |
| Db | 212     | gTTCtGCCCGGCAATCAGCAGGACAGGAATTAAGGGTATTATTAGGAACAGAG      | 153           |            |         |
|    |         |  |               |            |         |
| QY | 282     | catgtgcagtactcgtatttgccaatgtcctgcctgattattgaaaaacagcaggcac | 341           |            |         |
|    |         |  |               |            |         |
| Db | 152     | GATTTCAAAATGTCCCTGTTTGAGATGACATGCTATATCTAGAAAACCGCATTTGTC  | 93            |            |         |
|    |         |  |               |            |         |
| QY | 342     | agaataaaac   | 351           |            |         |
|    |         |  |               |            |         |
| Db | 92      | TCAGTCCAAC   | 83            |            |         |

|            |  |
|------------|--|
| RESULT     | 5  |
| AI942816   |  |
| LOCUS      | AI942816   |
| DEFINITION | f6c7b05.xl zebrafish WashU MPIMG EST Danio rerio CDNA 3' , mRNA          |
| ACCESSION  | AI942816   |
| VERSION    | AI942816.1   |
| KEYWORDS   | EST.   |
| SOURCE     | zebrafish.   |
| ORGANISM   | Danio rerio  |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;        |
|            | Actinopterygii; Neopercyiformes; Cyprinodontiformes; Cyprinidae;         |
|            | Cypriniformes; Cyprinodontiformes; Teleostei; Euteleostei; Osteichthyes; |
|            | Rasbora; Rasborinae; Danio.  |

embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

|            |         |      |      |       |
|------------|---------|------|------|-------|
| BASE COUNT | 154 a   | 60 c | 59 g | 187 t |
| ORIGIN     | CONCOT. |      |      |       |

|                       |       |                |         |             |
|-----------------------|-------|----------------|---------|-------------|
| Query Match           | 9.3%  | Score 35.4;    | DB 104; | Length 460; |
| Best Local Similarity | 50.9% | Pred. No. 3.5; |         |             |

195 tattaataattatccctgttadaatgtttaagaacagggcattcagtcggacaataac 254  
QY ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db TTTT|||||||TACAAATTGAAGTGCTTTAACAAAAACAATTAATTGTCCAAAAA 64

255 gaacccttgattctttggttgacctgtcatgtactgtatttgcgaatgtcctg 314  
QY ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db AACCTCATAATTATGTTCTTTAAAGTCCTATTAAATCAAGTAGTTTTGAACAAACACGACAG 124

315 cctgattatttgaataatcagcaggacaggaataaaactgtggaat 359  
QY ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db TATAAATTTTGAATTCATCTAGTGAAAATTATTATCATGTGGCAT 169

| RESULT     | 6  |
|------------|--|
| BE603735   |  |
| LOCUS      | 955 bp mRNA  |
| DEFINITION | HVSDNA0087003f Hordeum vulgare 5-45 DAP spike EST library                              |
| KEYWORDS   | HVSDNA0087003f (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0087003f, mRNA sequence. |
| ACCESSION  | BE603735   |
| VERSION    | BE603735.2   |
| KEYWORDS   | GI:13187835  |
| SOURCE     | EST.   |
| ORGANISM   | barley.  |
|            | Hordeum vulgare  |

BASE COUNT  
ORIGIN





```

QY 222 gtaagaacaggcattcagtcggacaaatagcaaccccttagttatcttggttcac 281
    ||| | ||||| ||||| ||| | | | | | | | | | | | | | | | | |
Db 334 GTCTGCGCAGGCAATCAGTAGGAGAAATAAACCAGTATTCATAGGAAAGAG 393
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 282 catgtcgtactcgtatttgcaatgctcctgctgatttgaataacagcagggac 341
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 GAAGTCAAAATGTCCCTGTTTGCAGATGACATGATTGTATATTAGAAAACCCCATTTGC 453

RESULT 9
AQ418525 AQ418525 487 bp DNA GSS 23-MAR-1999
LOCUS RPCI-11-20318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-20318,
DEFINITION DNA sequence.
ACCESSION AQ418525
VERSION AQ418525.1 GI:4476249
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 487)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished (1997)
Other_GSSs: RPCI-11-20318.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES
Source Location/Qualifiers
1..487
/organism="Homo sapiens"
/db_xref="GDB:757767"
/db_xref="taxon:9606"
/clone="RPCI-11-20318"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 165 a 102 c 96 g 124 t
ORIGIN

Query Match 9.1%; Score 34.6; DB 228; Length 487;
Best Local Similarity 54.3%; Pred. No. 6.2;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 208 cccctttagaatgtgttaagaacagggcattcagtcggacaaatagcaaccccttagtt 267
    ||| ||||| ||| | | | | | | | | | | | | | | | |
Db 127 CTCTGGTAGAATTCGGTCCGTCAGGCAATCAGGCGAGGAGAAATAAAGGGTCTTC 186
    ||| ||||| ||| | | | | | | | | | | | | | | | |

QY 268 atcttgggttcacatgctatctgatttgcgaatgctcctgctgattattga 327
    ||| ||||| ||| | | | | | | | | | | | | | | | |
Db 187 AATTAGGAAAAGAGGAAGTCAAAATGTCCCTGTTTGCAGATGACATGATTGTGTATCTAC 246
    ||| ||||| ||| | | | | | | | | | | | | | | | |

QY 328 aaatcagca 336
    ||| | | |

```

```

Db 247 AAAACCCCA 255

RESULT 10
AQ417598 AQ417598 555 bp DNA GSS 23-MAR-1999
LOCUS RPCI-11-203C8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203C8,
DEFINITION DNA sequence.
ACCESSION AQ417598
VERSION AQ417598.1 GI:4475441
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 555)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished (1997)
Other_GSSs: RPCI-11-203C8.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES
Source Location/Qualifiers
1..555
/organism="Homo sapiens"
/db_xref="GDB:7577623"
/db_xref="taxon:9606"
/clone="RPCI-11-203C8"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 192 a 114 c 110 g 139 t
ORIGIN

Query Match 9.1%; Score 34.6; DB 228; Length 555;
Best Local Similarity 54.3%; Pred. No. 6.3;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 208 cccctttagaatgtgttaagaacagggcattcagtcggacaaatagcaaccccttagtt 267
    ||| ||||| ||| | | | | | | | | | | | | | | | |
Db 150 CTCTGGTAGAATTCGGTCCGTCAGGCAATCAGGCGAGGAGAAATAAAGGGTCTTC 209
    ||| ||||| ||| | | | | | | | | | | | | | | | |

QY 268 atcttgggttcacatgctatctgatttgcgaatgctcctgctgattattga 327
    ||| ||||| ||| | | | | | | | | | | | | | | | |
Db 210 AATTAGGAAAAGAGGAAGTCAAAATGTCCCTGTTTGCAGATGACATGATTGTGTATCTAC 269
    ||| ||||| ||| | | | | | | | | | | | | | | | |

QY 328 aaatcagca 336
    ||| | | |
Db 270 AAAACCCCA 278

RESULT 11
CNS016H0 CNS016H0 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION

```



20y 150 ccaaaacgaaatttaactcctcgatgtaccagggtgtaatttagagatatttaaattatqcc 209

us-09-462-955-1\_copy\_611\_991.rst

Thu Oct 25 13:08:25 2001

Search completed: October 24, 2001, 13:13:47  
Job time: 11812 sec

. . . . .

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:26 ; Search time 180.6 Seconds  
(without alignments)  
399.377 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_611\_991

Perfect score: 381

Sequence: 1 cagcgacgacgcacacatc.....ttaaagtatgtcatctaa 381

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description      |
|------------|-------|-------------|--------|----|------------------|
| 1          | 112.4 | 29.5        | 1096   | 2  | US-08-418-071-4  |
| 2          | 110.8 | 29.1        | 1091   | 2  | US-08-418-071-5  |
| 3          | 89.6  | 23.5        | 1106   | 2  | US-08-418-071-3  |
| 4          | 88.6  | 23.3        | 1017   | 4  | US-08-793-634B-6 |
| 5          | 88.2  | 23.1        | 1022   | 4  | US-08-793-634B-2 |
| 6          | 62.8  | 16.5        | 1110   | 1  | US-08-202-186-14 |
| 7          | 59.6  | 15.6        | 982    | 3  | US-08-973-068-28 |
| 8          | 59.6  | 15.6        | 1110   | 1  | US-08-202-186-11 |
| 9          | 59.6  | 15.6        | 1111   | 1  | US-08-202-186-9  |
| 10         | 59.6  | 15.6        | 1111   | 1  | US-08-202-186-12 |
| 11         | 59.6  | 15.6        | 1111   | 1  | US-08-202-186-24 |
| 12         | 58.6  | 15.4        | 1103   | 1  | US-08-202-186-16 |
| 13         | 58.6  | 15.4        | 1105   | 1  | US-08-202-186-13 |
| 14         | 58    | 15.2        | 1109   | 1  | US-08-202-186-13 |
| 15         | 58    | 15.2        | 1111   | 1  | US-08-202-186-15 |
| 16         | 57    | 15.0        | 1104   | 1  | US-08-202-186-17 |
| 17         | 56.4  | 14.8        | 1111   | 1  | US-08-202-186-10 |
| 18         | 40.2  | 10.6        | 300    | 2  | US-08-418-071-2  |
| 19         | 33.8  | 8.9         | 287    | 2  | US-08-418-071-1  |
| 20         | 29.4  | 7.7         | 8920   | 4  | US-08-446-855A-1 |
| 21         | 29.4  | 7.7         | 8920   | 4  | US-09-150-741-1  |
| 22         | 28.8  | 7.6         | 10607  | 1  | US-08-078-090-3  |
| 23         | 28.4  | 7.5         | 1380   | 4  | US-09-499-505-4  |
| 24         | 28.4  | 7.5         | 1430   | 4  | US-09-499-505-3  |
| 25         | 28    | 7.3         | 28958  | 1  | US-08-258-261B-6 |
| 26         | 28    | 7.3         | 28958  | 1  | US-08-456-837-6  |
| 27         | 28    | 7.3         | 28958  | 1  | US-08-457-342-6  |

|    |      |     |       |   |                   |                    |
|----|------|-----|-------|---|-------------------|--------------------|
| 28 | 28   | 7.3 | 28958 | 1 | US-08-457-646A-6  | Sequence 6, Appli  |
| 29 | 28   | 7.3 | 28958 | 1 | US-08-458-076A-6  | Sequence 6, Appli  |
| 30 | 28   | 7.3 | 28958 | 1 | US-08-764-233A-4  | Sequence 4, Appli  |
| 31 | 28   | 7.3 | 28958 | 1 | US-08-764-233A-4  | Sequence 4, Appli  |
| 32 | 28   | 7.3 | 28958 | 1 | US-08-457-335A-6  | Sequence 6, Appli  |
| 33 | 28   | 7.3 | 28958 | 1 | US-08-729-214-6   | Sequence 6, Appli  |
| 34 | 28   | 7.3 | 28958 | 3 | US-09-028-934-6   | Sequence 6, Appli  |
| 35 | 28   | 7.3 | 49377 | 1 | US-08-764-233A-1  | Sequence 1, Appli  |
| 36 | 27.8 | 7.3 | 2570  | 2 | US-09-056-075-2   | Sequence 2, Appli  |
| 37 | 27.6 | 7.2 | 2899  | 2 | US-08-624-581-2   | Sequence 2, Appli  |
| 38 | 27.6 | 7.2 | 2917  | 2 | US-08-624-581-3   | Sequence 3, Appli  |
| 39 | 27.6 | 7.2 | 2959  | 2 | US-08-624-581-1   | Sequence 1, Appli  |
| 40 | 27.4 | 7.2 | 6476  | 4 | US-09-127-670-5   | Sequence 5, Appli  |
| 41 | 27.4 | 7.2 | 840   | 5 | PCT-US91-08177-12 | Sequence 12, Appli |
| 42 | 27.4 | 7.2 | 2256  | 6 | 5220013-1         | Patent No. 5220013 |
| 43 | 27.2 | 7.1 | 7323  | 5 | PCT-US91-08177-1  | Sequence 1, Appli  |
| 44 | 27.2 | 7.1 | 5892  | 3 | US-08-755-587-27  | Sequence 27, Appli |
| 45 | 27.2 | 7.1 | 7240  | 3 | US-08-755-587-15  | Sequence 15, Appli |
|    |      | 7.1 | 11283 | 2 | US-08-603-753D-3  | Sequence 3, Appli  |

## ALIGNMENTS

RESULT 1  
US-08-418-071-4  
; Sequence 4, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh  
; APPLICANT: You, Li-Ru  
; APPLICANT: Soong, Tai-Seng  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE  
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/418,071  
; FILING DATE: 06-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: DCB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1096 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; MOLECULE TYPE: genomic DNA  
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Banana Bunchy Top Virus (BBTV)  
US-08-418-071-4

```

Query Match      29.5%; Score 112.4; DB 2; Length 1096;
Best Local Similarity 57.9%; Pred. No. 6.1e-29;
Matches 219; Conservative 0; Mismatches 156; Indels 3; Gaps 1;

QY 4 cggacgacgcacacactcctcgtggtatcgacgacgagacgagagcggaagtccgtgt 63
DB 564 CTGATGTCGAAGTATCATCTGGGTATACGGTCCCAACGGAGGCGAAGGAAAGTCAACCT 623

QY 64 ttgccaaatatctcggactcagcccgactgtttctacacatgtggtgggaaccagaaagg 123
DB 624 TCGCAAGATATCTATCATTAATAACCCGGATCGGGATATATCAACGGTGGAAAGACGTCGG 683

QY 124 acgtattgtaccagtacatcgacgagaccccaacgaaatttaactcctcgatgtaccaggt 183
DB 684 ATATGATGCACATCATACGATGATCTCTGATATCATTTGGATTGATATATCCCGCGAA 743

QY 184 gtaatttagagatttaaatatgcctcgtttagaaatgttttaagaaacagggcattcagtt 243
DB 744 GTCAATTCAGATTATCTGAATTATGCGGTATAGACAAATTAAGATAGATTTTAATAA 803

QY 244 cggacaaatacgaacc---ccttagttatcttgggttcgaccatgtgcatgtactcgtat 300
DB 804 ATACAAATATACGAACCATCTGTGATTAGAAAAGATGGACAAAATCTCCATGTAATTGTTA 863

QY 301 ttgccaatgtcctcgtcgtatatttgaataatcagcagggacagaaataaactgtggaata 360
DB 864 TGGCAAAATGTGTCCTGATTATTTGAAATTTTCAGAAGATAGATAAATAATAATTAAAT 923

QY 361 tttaaaagtatgtgtcatc 378
DB 924 GTTGAGAAAGGAAACTTC 941

```

RESULT 2

```

US-08-418-071-5
; Sequence 5, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTI
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1091 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: circular
; MOLECULE TYPE: subgenomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2)"
; HYPOTHEITICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; US-08-418-071-5

Query Match      29.1%; Score 110.8; DB 2; Length 1091;
Best Local Similarity 57.7%; Pred. No. 2.1e-28;
Matches 218; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

QY 4 cggacgacgcacacactcctcgtggtatcgacgagacgagagcggaagtccgtgt 63
DB 559 CTGATGTCGAAGTATCATCTGGGTATACGGTCTCCCGAGGCGAAGGAAAGTCAACCT 618

QY 64 ttgccaaatatctcggactcagcccgactgtttctacacatgtggtgggaaccagaaagg 123
DB 619 TCGCAAGATATCTATCATTAATAACCTGGATGATATCAACGGTGGAAAGACGTCGG 678

QY 124 acgtattgtaccagtacatcgacgagaccccaacgaaatttaactcctcgatgtaccaggt 183
DB 679 ATATGATGCACATCATACGATGATCTCTGATATCATTTGGATTGATATATCCCGCGAA 738

QY 184 gtaatttagagatttaaatatgcctcgtttagaaatgttttaagaaacagggcattcagtt 243
DB 739 GTCAATTCAGATTATCTGAATTATGCGGTATAGACAAATTAAGATAGATTTTAATAA 798

QY 244 cggacaaatacgaacc---ccttagttatcttgggttcgaccatgtgcatgtactcgtat 300
DB 799 ATACAAATATACGAACCATCTGTGATTAGAAAAGATGGACAAAATGTCCATGTAATTGTTA 858

QY 301 ttgccaatgtcctcgtcgtatatttgaataatcagcagggacagaaataaactgtggaata 360
DB 859 TGGCAAAATGTGTCCTGATTATTTGAAATTTTCAGAAGATAGATAAATAATAATTAAAT 918

QY 361 tttaaaagtatgtgtcatc 378
DB 919 GTTGAGAAAGGAAACTTC 936

```

RESULT 3

```

US-08-418-071-3
; Sequence 3, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.

```

```

; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; US-08-418-071-3

```

```

Query Match      23.5%; Score 89.6; DB 2; Length 1106;
Best Local Similarity 52.7%; Pred. No. 3.8e-21;
Matches 194; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

```

```

QY 2 agcggacgatgcacaaatcctctggatcgagacgagagcggagagcggagtcctcg 61
Db 559 ACCGGATGATCGAAGCATCATCTGGGTCTATGGTCTTATGGTAAATCAGGGTAAATCAAC 618
QY 62 gttggcaaatatcggactcaagcccgactgttctacacatggttggaaccagaaa 121
Db 619 ATATGCGAAGTCACTAATCAAGAAGGATTTGGTCTACACGAGGGTGGGAAGAGAGAA 678
QY 122 ggaactgttaccagtcacatcgagagcaccacaaacgaaatttaattcctcgatgacc 181
Db 679 TATCTTTATCTCTATGTGACGAAAGGATCTGACAGCATATAGTATTTCATATTCCTCG 738
QY 182 gtgaatttagagatttaaatatgccctgttgagaatgtgttaagaacagggttcag 241
Db 739 TTGTAATCAGGATTTAAATATATGTAATAGAGGCATTAAAGGATAGGGTTATAGA 798
QY 242 ttcggacaaatacgaaccccttagttatcttgggttcgaccatgtgcactgactgtatt 301
Db 799 GAGTACTAAATACAAACCCATAAGATAGTTGAAATAGGTAAATACATGTAATCGTAT 858
QY 302 tgcaatgtcctcctgattatttgaaatcagcaggagacagaataaaaactgtggaatat 361
Db 859 GCGCAATTTTCATGCTGCTGCTGTAATAATCTCCGAAGATCGAATAAATAATCATTTATTG 918
QY 362 ttaaaata 369
Db 919 CTGAAGAA 926

```

```

RESULT 4
US-08-793-634B-6
; Sequence 6, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.
; APPLICANT: Larkin, Philip J.
; APPLICANT: Taylor, William C.
; APPLICANT: Marchall, Jerry S.
; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City

```

```

; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,634B
; FILING DATE: June 9, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-793-634B-6

```

```

Query Match      23.3%; Score 88.6; DB 4; Length 1017;
Best Local Similarity 54.1%; Pred. No. 8e-21;
Matches 203; Conservative 0; Mismatches 169; Indels 3; Gaps 1;

```

```

QY 2 agcggacgatgcacaaatcctctggatcgagacgagagcggagagcggagtcctcg 61
Db 539 ACCAGATTATAGACGATAATCTGGTCTATGGACCTCTGCTGTAATGAAGCAATCTAC 598
QY 62 gttggcaaatatcggactcaagcccgactgttctacacatggttggaaccagaaa 121
Db 539 ATTTGCAAGACATCTGTCTATTGAAAGATGGTTGGGGTTTATCTGCTGGAGGAAGACACA 658
QY 122 ggaactgttaccagtcacatcgagagcaccacaaacgaaatttaattcctcgatgacc 181
Db 659 AGATATGATGATCATCTGTGACTGCTGAGCCCTAAGAAATAATTTGGGTATTGACATACCCAG 718
QY 182 gtgtaatttagagatttaaatatgccctgttgagaatgtgttaagaacagggttcag 241
Db 719 AGTTAGTTTACAGAGTATGTAATTTATGGTGTAAATAGAACAGGTTAAGAAATAGGGTAATGGT 778
QY 242 ttcggacaaatacgaaccccttagttatcttgggttcgaccatgtgcactgactgt 298
Db 779 GAATACTAAGTATGAGCATGTGTAATGCGGGATGATAATCATCTGTTTATGTAATTTGT 838
QY 299 atttgcaaatgtcctgctgattatttgaaatcagcaggagacagaataaaaactgtgga 358
Db 839 GTTTGCAAAATGCTACTCCAGATTTGGGAAATAATTAAGTCAAGATACATAAATAAATTAATTCG 898
QY 359 tatttaagaatgtg 373
Db 899 TTGTTGAAAACTCTG 913

```

```

RESULT 5
US-08-793-634B-2
; Sequence 2, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.

```

|                       |        |                    |    |
|-----------------------|--------|--------------------|----|
| Best Local Similarity | 53.28; | Pred. NO. 3.5e-12; | 2; |
| Matches               | 183.   | Conservative       | 0: |
| Mismatches            | 152:   | Indels             | 9: |
| Gaps                  | 9:     | Gaps               | 2: |

2 agcgacgatacgacaatcctctggatatcgggacgagacggaggagacgggaagtcct 61

|                           |        |                    |            |
|---------------------------|--------|--------------------|------------|
| Best Local Similarity     | 54.58; | Pred. No. 1.1e-20; |            |
| Matches 100. Conservative | 0;     | Mismatches 163;    | Indels     |
|                           |        |                    | 3: Gaps 1: |

## RESULT 6

RESULT: 6  
US-08-202-186-14

US-08-973-068-28  
; Sequence 28, Application US/08973068



```

RESULT      8
US-08-202-186-11
; Sequence 11, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT:  KARAN, Mirko
; APPLICANT:  BURNS, Thomas M.
; APPLICANT:  DALE, James L.
; APPLICANT:  HARDING, Robert M.
; TITLE OF INVENTION:  DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES:  60
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE:  Foley & Lardner
; STREET:     3000 K Street, N.W.
; CITY:       Washington, D.C.
; COUNTRY:    USA
; ZIP:        20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:     IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS

```

```

RESULT      9
US-08-2002-186-9
; Sequence 9, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.25

```

Query Match 15.6%; Score 59.6; DB 1; Length 1111;  
Best Local Similarity 52.6%; Pred. No. 6.8e-11;  
Matches 181: Conservative 0; Mismatches 154; Indels 9

[illegible]

```

RESULT 11
US-08-202-186-24
; Sequence 24, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186

```

|                           |        |                    |           |              |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match               | 15.6%  | Score 59.6;        | DB 1;     | Length 1111; |
| Best Local Similarity     | 52.6%; | Pred. No. 6.8e-11; |           |              |
| Matches 181; Conservative | 0;     | Mismatches 154;    | Indels 9; | Gaps 2;      |

10 atgcacaaatcctctggatatacgcagacagacagacagacagcgggaagtcggtttggca 69  
 616 ATCGGAGAAATAATTTGGGTCTATGGCCCAATGAGGAGAGAGAAAGACAACGTA TGCA 675  
 70 aatatctcggactcaagccgcactggttctacacatggtggaaccagaaagacgctat 129  
 676 AACATCTAATGAAGACGAGAAATGCGTTTTATTCTCAGGAGGAGAAATCATTTGGATATAT 735  
 130 tgtaccagtcacatcgcaggaccacaaacgaaataattaaacctcgatgaccacaggtgtaatt 189  
 736 GTAGACTGCTATAATACGAGGATAT-----TGTATATTGTGATATCCAGATGCAAG 789  
 190 tagagttattcaaatatgcctcttagaattgtttaagaacagggcattcagttoggaca 249  
 790 AGGATTATTAAATATTGGTATTATAGAGGAATTTAAGAAATGCAATAATTTCAAGCGGGA 849  
 250 aatcagaaccccttagttatcttggttgcagcattgcatctactgctatttggccaatg 309  
 850 AATTATGAACCCGTTTGAAGATAG---TAGAATATGTCGAAGTCATTGTAATGCGTAACT 906  
 310 tctgcgcctgattttgaaataatcagcaggacagaataaaaaactg 353  
 907 TCCCTTCGGAAGGAAGGAATCTTTTCTGAAGATCGGAATAAAGATTG 950

```

RESULT 10
; JS-08-202-186-12
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```



APPLICATION NUMBER: US/08/202,186  
FILING DATE: 24-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: JEFFERY, Donald D.  
REGISTRATION NUMBER: 19,980  
REFERENCE/DOCKET NUMBER: 71611/102 FKE  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

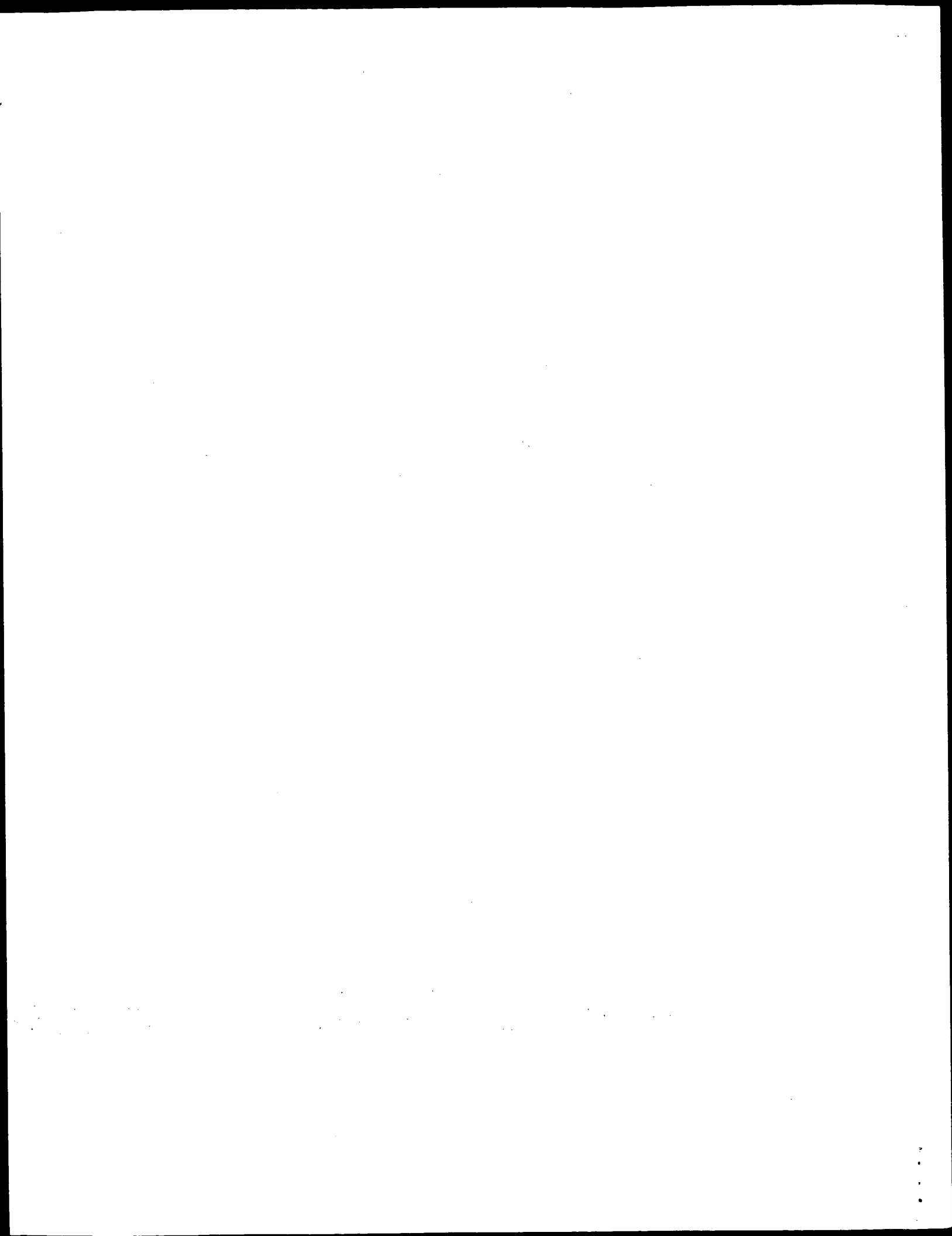
|     | Query Match   | 15.4%;       | Score 58.6;        | DB 1;           | Length 1105;      |
|-----|---|--------------|--------------------|-----------------|-------------------|
|     | Best Local Similarity   | 52.5%;       | Pred. No. 1.5e-10; |                 |                   |
|     | Matches 180;  | Conservative | 0;                 | Mismatches 154; | Indels 9; Gaps 2; |
| 20y | 10 atcgacaaatccctctggtatgcgagcagagcagagcagcggaagtcctggtttgcc      | 69           |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 618 ACCGAAGGATAATTTGGGCTATGGCCCAATGGAGGAGAAAGACAACTTATGCAA        | 677          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 70 aatctcggactcaagcccgactggtctacacatggtgtgacccagaaggagcgat        | 129          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 678 AATATTAAATGAACACGAAGAATCGCTTTATTTCGCCAGGAGAAATCATTTGGATAT     | 737          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 130 tgtaccagtcacatcgaggaccacaaacgaatattacatcctgatgtaccaggtgaatt   | 189          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 738 GTAGATTGTATAATTTATGAGGAATA-----GTTATATTGTATATCCCATGCAAG       | 791          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 190 tagagatttaaatatgcctgttagaatggtttaagaacacaggcattcagttcggaca    | 249          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 792 AGGAATATTTAAACTATCGGTTATTAGAGAATTTTAAAGATGGAAATTTCAACGGGGA    | 851          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 250 aatcagaaccccttagttatctctgggtctcgaccatgtgcattgactcgattttgccatg | 309          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 852 ATATATGAACCCGTTTGGAAATTG---TAGAATATGTGGNAGTCATTGTAAATGGCTAACT | 908          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 310 tctctgcctgattatttgaataatcagcagggacagataaaact                  | 352          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |
| 20y | 909 TCCTTCCGAAGGAAGGAATCTTTTCAGAAGATCGAATAAAGCT                   | 951          |                    |                 |                   |
| 20y |   |              |                    |                 |                   |

```

RESULT      14
US-08-202-186-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
;
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:24 ; Search time 393.52 Seconds  
(without alignments)  
448.364 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_711\_991

Perfect score: 281

Sequence: 1 cacatgtgtggaacagaa.....ttaaagtatgtcatctaa 281

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

```
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 70.8  | 25.2        | 1091   | 20 | AAV71834    |
| 2          | 70.8  | 25.2        | 1091   | 21 | AAA38947    |
| 3          | 70.8  | 25.2        | 1096   | 20 | AAV71833    |
| 4          | 70.8  | 25.2        | 1096   | 21 | AAA38950    |
| 5          | 70.8  | 25.2        | 1106   | 20 | AAV71832    |
| 6          | 70.6  | 25.1        | 886    | 20 | AAV71837    |
| 7          | 69.2  | 24.6        | 1106   | 21 | AAA38946    |
| 8          | 67.4  | 24.0        | 593    | 20 | AAA34686    |
| 9          | 65.2  | 23.2        | 1017   | 17 | AAAT13165   |
| 10         | 61.8  | 22.0        | 1022   | 17 | AAAT13161   |
| 11         | 44.6  | 15.9        | 1105   | 19 | AAV24093    |

|    |      |      |       |    |          |                     |
|----|------|------|-------|----|----------|---------------------|
| 12 | 44.4 | 15.8 | 1103  | 19 | AAV24091 | Banana bunchy top   |
| 13 | 43.8 | 15.6 | 936   | 22 | AAF58252 | Oligonucleotide D1  |
| 14 | 43.8 | 15.6 | 936   | 22 | AAF58254 | Oligonucleotide D1  |
| 15 | 43.8 | 15.6 | 936   | 22 | AAF58257 | Oligonucleotide D1  |
| 16 | 43.8 | 15.6 | 936   | 22 | AAF58259 | Oligonucleotide D2  |
| 17 | 43.8 | 15.6 | 936   | 22 | AAF58262 | Oligonucleotide D2  |
| 18 | 43.8 | 15.6 | 938   | 22 | AAF58255 | Oligonucleotide D1  |
| 19 | 43   | 15.3 | 1110  | 19 | AAV24089 | Banana bunchy top   |
| 20 | 43   | 15.3 | 1111  | 19 | AAV24087 | Banana bunchy top   |
| 21 | 42.8 | 15.2 | 1104  | 19 | AAV24092 | Banana bunchy top   |
| 22 | 41.8 | 14.9 | 936   | 22 | AAF58252 | Oligonucleotide D1  |
| 23 | 41.8 | 14.9 | 936   | 22 | AAF58254 | Oligonucleotide D1  |
| 24 | 41.8 | 14.9 | 936   | 22 | AAF58257 | Oligonucleotide D1  |
| 25 | 41.8 | 14.9 | 936   | 22 | AAF58259 | Oligonucleotide D2  |
| 26 | 41.8 | 14.9 | 936   | 22 | AAF58262 | Oligonucleotide D2  |
| 27 | 41.8 | 14.9 | 938   | 22 | AAF58255 | Oligonucleotide D1  |
| 28 | 41.4 | 14.7 | 982   | 18 | AAV49405 | Banana bunchy top   |
| 29 | 41.4 | 14.7 | 1109  | 19 | AAV24088 | Banana bunchy top   |
| 30 | 41.4 | 14.7 | 1111  | 19 | AAV24077 | Banana bunchy top   |
| 31 | 41.4 | 14.7 | 1111  | 19 | AAV24084 | Banana bunchy top   |
| 32 | 40   | 14.2 | 1111  | 19 | AAV24085 | Banana bunchy top   |
| 33 | 39.8 | 14.2 | 1110  | 19 | AAV24086 | Banana bunchy top   |
| 34 | 39.8 | 14.2 | 1111  | 19 | AAV24090 | Banana bunchy top   |
| 35 | 37   | 13.2 | 244   | 22 | AAF58238 | Oligonucleotide D1  |
| 36 | 34.2 | 12.2 | 244   | 22 | AAF58238 | Oligonucleotide D1  |
| 37 | 32.6 | 11.6 | 5053  | 19 | AAV22816 | Nucleotide sequence |
| 38 | 32.6 | 11.6 | 5055  | 19 | AAV17236 | DNA from a region   |
| 39 | 31.4 | 11.2 | 41599 | 21 | AAZ35351 | Cosmid including s  |
| 40 | 30.8 | 11.0 | 41599 | 21 | AAZ35351 | Cosmid including s  |
| 41 | 30.6 | 10.9 | 9345  | 21 | AAZ93966 | Mouse uromodulin p  |
| 42 | 29.8 | 10.6 | 1510  | 17 | AAV29628 | DNA mismatch repai  |
| 43 | 29.6 | 10.5 | 811   | 21 | AAQ50650 | Arabidopsis thalia  |
| 44 | 29.4 | 10.5 | 8920  | 15 | AAQ62924 | Carbamoyl-phosphat  |
| 45 | 29.2 | 10.4 | 148   | 21 | AAA41338 | Human secreted exp  |

## ALIGNMENTS

RESULT 1  
AAV71834 standard; DNA; 1091 BP.  
XX AAV71834;  
XX AAV71834;  
XX 10-FEB-1999 (first entry)  
XX BBTV DNA II clone (2) nucleotide sequence.

DE BBTV DNA II clone (2) nucleotide sequence.  
XX Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
XX Banana bunchy top virus.

XX US5846705-A.  
XX 08-DEC-1998.  
XX 06-APR-1995; 95US-0418071.  
XX 06-APR-1995; 95US-0418071.  
XX (BIOT-) DEV CENT BIOTECHNOLOGY.

XX Soong T, Wu R, You L;  
XX WPI; 1999-059037/05.

XX Nucleic acids having banana bunchy top virus component sequences -  
PT used to design primers for use in polymerase chain reaction  
PT detection of the virus  
XX

PS Disclosure; Fig 12A-C; 27pp; English.

XX This represents the nucleotide sequence of a banana bunchy top virus  
CC (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention  
CC provides nucleic acid sequences associated with BBTV that can be used in  
CC a PCR technique for detecting BBTV. The nucleic acid sequences (AAV71830  
CC to AAV71833) are used as the basis for the construction of PCR primers,  
CC to detect BBTV infection. The PCR technique is used for detecting BBTV  
CC in plant tissues (preferably banana, especially Musa species). The  
CC virus, one of the most important banana species viruses, causes phloem  
CC damage and is transmitted by aphids. PCR detection gives accurate,  
CC reliable and specific determination of absence or presence of the  
CC virus.  
XX Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other;  
SQ

Query Match 25.2%; Score 70.8; DB 20; Length 1091;  
Best Local Similarity 56.2%; Pred. No. 4.1e-13;  
Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;  
XX

QY 8 ggtggaaccagaaagcgtattgtaccagtcacatcgaggaccacaaacgaaatttaac 67  
DB 663 ggtggaagcgtcggtatgatgcacatcatacagatggtatcctgataatcattgatt 722  
QY 68 ctccgtatccagcgtgtaatttagagttatttaattatgcctgttagaattgtttaag 127  
DB 723 attgatacccccagagcattcagattatctgaattatggttattagacacaaattaag 782  
QY 128 aacaggcgttcagttcgacacaaatacgaacc---ccttagttatcttggttcaccat 184  
DB 783 aatagagttttaataataacaaatacgaaccatgtgtgattagaaaagatggacaaaat 842  
QY 185 gtccatgtactcgtatttgcgaatgtccctgctgatttattgaaatcagcaggacaga 244  
DB 843 gtccatgtattgttaggcaaatgtgtgctgatttattgaaatcagcaggacaga 902  
QY 245 ataaactgtggaattatttaagtagtgtgtcatc 278  
DB 903 ataaaaataattattgttgagaaggaaacttc 936

RESULT 2  
AAA38947  
ID AAA38947 standard; DNA; 1091 BP.  
XX  
AC AAA38947;  
XX  
DT 23-AUG-2000 (first entry)  
XX  
DE Banana bunchy top virus related nucleotide sequence #2.  
KW Banana bunchy top virus; BBTV; detection; ds.  
XX  
OS Banana bunchy top virus.  
XX  
PN TW360710-A.  
XX  
PD 11-JUN-1999.  
XX  
PF 30-JUN-1994; 94TW-0106105.  
XX  
PR 30-JUN-1994; 94TW-0106105.  
XX  
PA (BIOT-) DEV CENT BIOTECHNOLOGY.  
XX  
PI Wu R, You L, Song T;  
XX  
DR WPI; 2000-316145/27.  
XX  
XX Two circular single-stranded DNAs associated with banana bunchy top  
PT virus and detection of the virus -  
XX  
XX Claim 2; Page 1; 7pp; Chinese.

XX The present invention describes two circular single-stranded DNAs  
CC associated with banana bunchy top virus (BBTV) and the encoded proteins.  
CC Also described is a method of detecting the BBTV virus using the  
CC polymerase chain reaction (PCR). The present sequence represents a  
CC BBTV related nucleotide sequence from the present invention.  
XX  
SQ Sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;  
XX

Query Match 25.2%; Score 70.8; DB 21; Length 1091;  
Best Local Similarity 56.2%; Pred. No. 4.1e-13;  
Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;  
XX

QY 8 ggtggaaccagaaagcgtattgtaccagtcacatcgaggaccacaaacgaaatttaac 67  
DB 663 ggtggaagcgtcggtatgatgcacatcatacagatggtatcctgataatcattgatt 722  
QY 68 ctccgtatccagcgtgtaatttagagttatttaattatgcctgttagaattgtttaag 127  
DB 723 attgatacccccagagcattcagattatctgaattatggttattagacacaaattaag 782  
QY 128 aacaggcgttcagttcgacacaaatacgaacc---ccttagttatcttggttcaccat 184  
DB 783 aatagagttttaataataacaaatacgaaccatgtgtgattagaaaagatggacaaaat 842  
QY 185 gtccatgtactcgtatttgcgaatgtccctgctgatttattgaaatcagcaggacaga 244  
DB 843 gtccatgtattgttaggcaaatgtgtgctgatttattgaaatcagcaggacaga 902  
QY 245 ataaactgtggaattatttaagtagtgtgtcatc 278  
DB 903 ataaaaataattattgttgagaaggaaacttc 936

RESULT 3  
AAV71833  
ID AAV71833 standard; DNA; 1096 BP.  
XX  
AC AAV71833;  
XX  
DT 10-FEB-1999 (first entry)  
XX  
DE BBTV DNA II clone (2-17) nucleotide sequence.  
XX  
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;  
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.  
XX  
OS Banana bunchy top virus.  
XX

Key Location/Qualifiers  
FH TATA\_signal 1..7  
FT /\*tag= a  
FT stem\_loop 8..38  
FT /\*tag= b  
FT CDS 70..928  
FT /\*tag= c  
FT /\*product= "ORF-V2 product"  
FT /\*transl\_except= (pos:215..217, aa:Gly)  
FT polyA\_signal 533..538  
FT /\*tag= d  
FT polyA\_signal 799..804  
FT /\*tag= e  
FT polyA\_signal 907..912  
FT /\*tag= f  
FT polyA\_signal 1030..1035  
FT /\*tag= ge  
XX  
XX US5846705-A.  
XX  
XX 08-DEC-1998.  
XX  
XX 06-APR-1995; 95US-0418071.  
XX









```

XX OS Subterranean clover stunt virus isolate F.
XX FH Key Location/Qualifiers
XX FT CDS 48..905
XX FT /*tag= a
XX WO9606932-Al.
XX PD 07-MAR-1996.
XX PF 30-AUG-1995; 95WO-AU00552.
XX PR 07-NOV-1994; 94AU-0009281.
XX PR 30-AUG-1994; 94AU-0007770.
XX PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX PA (AUSU ) UNIV AUSTRALIAN NAT.
XX PI Boevink PC, Chu PWG, Keese PK, Khan RL, Larkin PJ;
XX PI Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
XX DR WPI; 1996-160363/16.
XX CC Circovirus transcription regulatory sequences and related constructs
XX PT - useful in plants, esp. leguminous plants, for the modulation of
XX PT gene expression
XX PS Claim 4; Page 62; 121pp; English.
XX CC The subterranean clover stunt virus (SCSV) genome has at least 7
XX CC distinct ssDNA components, designated segments 1-7 (AAT13160-66), each
XX CC contg. 1 major open reading frame and a non-coding region. Segment
XX CC 6 is predicted to be a viral replication-associated protein gene.
XX CC Genetic constructs useful in the genetic engineering of plants (esp.
XX CC legumes) comprise 1 or more heterologous gene(s) operatively linked
XX CC to a promoter region, and in some cases also a terminator region,
XX CC selected from segments 1-7. The transcription regulators facilitate
XX CC expression of foreign genes in plants and also facilitate control of
XX CC levels of gene expression in different plant tissue types.
XX SQ Sequence 1017 BP; 312 A; 160 C; 265 G; 280 T; 0 other;

Query Match 23.2%; Score 65.2; DB 17; Length 1017;
Best Local Similarity 55.2%; Pred. No. 2, 5e-11;
Matches 149; Conservative 0; Mismatches 118; Indels 3; Gaps 1;

QY 7 tgggtgaaccagaaaggacgtattgtaccagtcacatcgaggaccacaaacgaaatttaatt 66
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 644 tggagaaagacacagatgatgatctgtgactgtgagcctgaagaataattgggt 703

QY 67 cctcgtatgccagtgtaattagagatttaattatgcctgtttagaattgttttaa 126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 704 atttgacataccagagtgattgagtgatgtaattatggtgaataagaacaggtttaa 763

QY 127 gaacagggaattcagttcggaacaaatacgaaccccttagttatcttgggtcgacca--- 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 764 gaatagggttaattggtgaataactaagatgatgacatgtgtaatgcgggatgataatcacc 823

QY 184 tgtgcgtactcgtatttgccaatgtcctgctgattatttgaaatacagcaggagacag 243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 824 ttttcattgaattgttttgcaaatgtactccagatttgggaaataatgaatgaagatag 883

QY 244 aataaactgtggaattatttaagtatgtg 273
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 884 aataaaataattcgtttgtgaaaactctg 913

RESULT 10
AAT13161
ID AAT13161 standard; DNA; 1022 BP.
XX

```

```

AC AAT13161;
XX 23-MAY-1996 (first entry)
XX SCSV segment 2.
XX SCSV; promoter; transcription; transgenic plant; legume;
XX KW gene expression; crop improvement; ss.
XX OS Subterranean clover stunt virus isolate F.
XX FH Key Location/Qualifiers
XX FT CDS 82..924
XX FT /*tag= a
XX PN WO9606932-Al.
XX PD 07-MAR-1996.
XX PF 30-AUG-1995; 95WO-AU00552.
XX PR 07-NOV-1994; 94AU-0009281.
XX PR 30-AUG-1994; 94AU-0007770.
XX PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX PA (AUSU ) UNIV AUSTRALIAN NAT.
XX PI Boevink PC, Chu PWG, Keese PK, Khan RL, Larkin PJ;
XX PI Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
XX DR WPI; 1996-160363/16.
XX CC Circovirus transcription regulatory sequences and related constructs
XX PT - useful in plants, esp. leguminous plants, for the modulation of
XX PT gene expression
XX PS Claim 4; Page 58; 121pp; English.
XX CC The subterranean clover stunt virus (SCSV) genome has at least 7
XX CC distinct ssDNA components, designated segments 1-7 (AAT13160-66), each
XX CC contg. 1 major open reading frame and a non-coding region. Segment
XX CC 2 is predicted to be a viral replication-associated protein gene.
XX CC Genetic constructs useful in the genetic engineering of plants (esp.
XX CC legumes) comprise 1 or more heterologous gene(s) operatively linked
XX CC to a promoter region, and in some cases also a terminator region,
XX CC selected from segments 1-7. The transcription regulators facilitate
XX CC expression of foreign genes in plants and also facilitate control of
XX CC levels of gene expression in different plant tissue types.
XX SQ Sequence 1022 BP; 309 A; 183 C; 245 G; 285 T; 0 other;

Query Match 22.0%; Score 61.8; DB 17; Length 1022;
Best Local Similarity 52.1%; Pred. No. 3, 1e-10;
Matches 138; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 2 acatgtgtgtgaaccagaaaggacgtattgtaccagtcacatcgaggaccacaaacgaaat 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 661 acagcggagggaagaccagcagctattatattgtatgtctcaagaccagagagagaaat 720

QY 62 ttaatcctcgtatgcaccaggtgaatttagagatttaattatccctttagaattgt 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 721 attgcgtttgatgtctccaggtgtcttcgagagatgatgaactatcaggcgatggagatg 780

QY 122 gttagaacagggcatttcagttcggacaataacgaaccccttagttatcttgggttcgac 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 781 ttgaagaacagagatttttgcaagtacaaataatagcctgtagattttgtattaggaag 840

QY 182 catgtgcattcgtactgtatttggccaatgtcctgctgattatttgaaatacagcaggagac 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 841 ttagtcttaattatgtttgtccaactgtggcaactgacccccacgcgcataagtggagac 900

QY 242 agaataaaactgtggaataattttaaa 266

```



```

XX PD 01-FEB-2001.
XX PF
XX PD 26-JUL-2000; 2000WO-US20476.
XX PR
XX PF 26-JUL-1999; 99US-0145695.
XX PR
XX PR 17-MAR-2000; 2000US-0190259.
XX PA
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI
XX PI Umek RM;
XX DR
XX DR WPI; 2001-159728/16.
XX PT
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT a single surface
XX PS
XX PS Example 6; Page 127; 159pp; English.
XX CC
XX CC The present invention relates to a composition comprising two nucleic
XX CC acids each containing an electron-transfer group (ETM) having
XX CC different redox potentials. The invention is used for electronic
XX CC detection of nucleic acids, especially of substitutions (mismatches)
XX CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX CC monitoring gene expression.
XX SQ
XX SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 15.6%; Score 43.8; DB 22; Length 936;
Best Local Similarity 2.4%; Pred. No. 0.00018;
Matches 6; Conservative 156; Mismatches 93; Indels 0; Gaps 0;

QY 27 tatgtaccagtcacatgcagagcccaaaacgaaatttaactctcgatgtaccaggtgta 86
Db 600 WWWWWWCCCCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 541
QY 87 attagagatttaattatgcctgttagaattgttaagaacaggcattcagttcgg 146
Db 540 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 481
QY 147 acaaatcacgaaccccttagtctatcttggttcgacacatgtgcatgtactcgattgcca 206
Db 480 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 421
QY 207 atgtcctgcctgattatttgaaatcacgagggacagataaaactgtggaattattaa 266
Db 420 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWW 361
QY 267 gtatgtgtcatctaa 281
Db 360 WWWWWWWWWWWW 346

RESULT 14
AAF58254/C
ID AAF58254 standard; DNA; 936 BP.
XX AC
XX AC AAF58254;
XX DT
XX DT 24-APR-2001 (first entry)
XX DE
XX DE Oligonucleotide D1875.
XX KW
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS
XX OS Synthetic.
XX PN
XX PN WO200107665-A2.
XX PD
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.

```

```

XX PF
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR
XX PR 26-JUL-1999; 99US-0145695.
XX PR
XX PR 17-MAR-2000; 2000US-0190259.
XX PA
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI
XX PI Umek RM;
XX DR
XX DR WPI; 2001-159728/16.
XX PT
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT a single surface
XX PS
XX PS Example 6; Page 127; 159pp; English.
XX CC
XX CC The present invention relates to a composition comprising two nucleic
XX CC acids each containing an electron-transfer group (ETM) having
XX CC different redox potentials. The invention is used for electronic
XX CC detection of nucleic acids, especially of substitutions (mismatches)
XX CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX CC monitoring gene expression.
XX SQ
XX SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

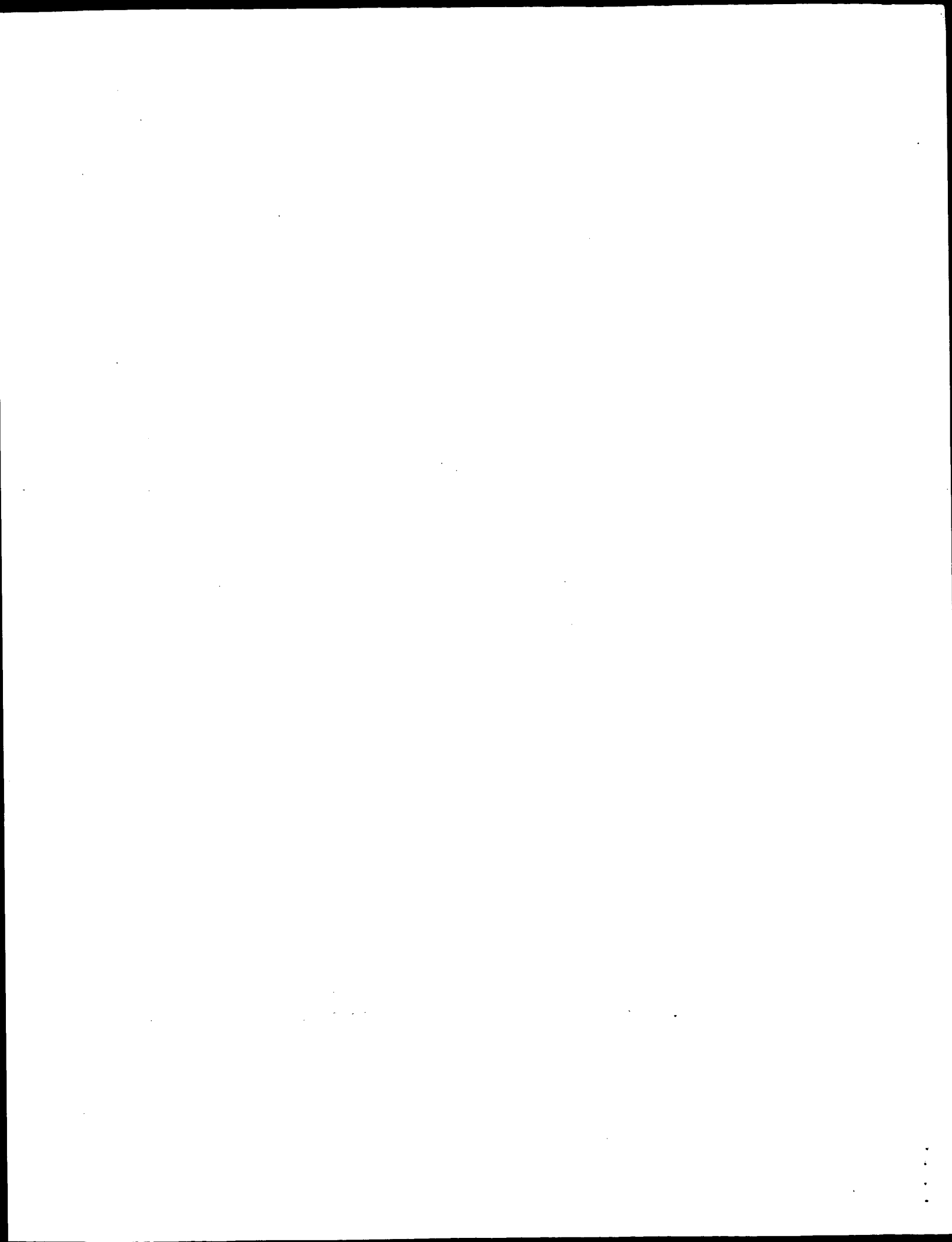
Query Match 15.6%; Score 43.8; DB 22; Length 936;
Best Local Similarity 2.4%; Pred. No. 0.00018;
Matches 6; Conservative 156; Mismatches 93; Indels 0; Gaps 0;

QY 27 tatgtaccagtcacatgcagagcccaaaacgaaatttaactctcgatgtaccaggtgta 86
Db 600 WWWWWWCCCCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 541
QY 87 attagagatttaattatgcctgttagaattgttaagaacaggcattcagttcgg 146
Db 540 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 481
QY 147 acaaatcacgaaccccttagtctatcttggttcgacacatgtgcatgtactcgattgcca 206
Db 480 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 421
QY 207 atgtcctgcctgattatttgaaatcacgagggacagataaaactgtggaattattaa 266
Db 420 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWW 361
QY 267 gtatgtgtcatctaa 281
Db 360 WWWWWWWWWWWW 346

RESULT 15
AAF58257/C
ID AAF58257 standard; DNA; 936 BP.
XX AC
XX AC AAF58257;
XX DT
XX DT 24-APR-2001 (first entry)
XX DE
XX DE Oligonucleotide D1954.
XX KW
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS
XX OS Synthetic.
XX PN
XX PN WO200107665-A2.
XX PD
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:47 ; Search time 5479.82 Seconds  
(without alignments)  
484.733 Million cell updates/sec

Title: us-09-462-955-1\_copy\_711\_991

Perfect score: 281

Sequence: 1 cacatgtgtgtgaaccagaa.....ttaaagtatgtgtatctaa 281

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*





|   |     |  |     |  |  |  |  |  |  |
|---|-----|--|-----|--|--|--|--|--|--|
| /db_xref="taxon:9606"<br>/clone_lib="STU277"<br>/dev_stage="Adult"<br>/note="Organ: stomach; Vector: puc18; Site_1: Sma1;<br>Site_2: Sma1; A mini-library was made by cloning products<br>derived from QUESPES PCR (U.S. Letters Patent application<br>No. 196,716 - Ludwig Institute for Cancer Research)<br>profiles into the pUC 18 vector. Reverse transcription of<br>tissue mRNA and cDNA amplification were performed under<br>low stringency conditions."<br>193 a 119 c 89 g 117 t |     |  |     |  |  |  |  |  |  |
| BASE COUNT<br>ORIGIN  |     |  |     |  |  |  |  |  |  |
| Query Match 13.1%; Score 36.8; DB 121; Length 518;<br>Best Local Similarity 61.5%; Pred. No. 1.7;<br>Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;   |     |  |     |  |  |  |  |  |  |
| QY  | 45  | aggaccgaagaaatttaactctgatgataccagggtgaatttagagattttaaat  | 104 |  |  |  |  |  |  |
| DB  | 168 | AGGATTAAACCAAAATTGATCTACATGGCACACAGCAAAAATATAGTAACAAGATT | 227 |  |  |  |  |  |  |
| QY  | 105 | atgcctgttagaattgtttaagacaggggcatcca                      | 140 |  |  |  |  |  |  |
| DB  | 228 | GTGCTCAGAAGTAACGCTTTAACACCTGCCCAATCA                     | 263 |  |  |  |  |  |  |
| RESULT 4  |     |  |     |  |  |  |  |  |  |
| AQ089903/c 318 bp DNA GSS 26-AUG-1998   |     |  |     |  |  |  |  |  |  |
| LOCUS HS_3001_A2_H04_MR_CIT Approved Human Genomic Sperm Library D Homo   |     |  |     |  |  |  |  |  |  |
| DEFINITION sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.  |     |  |     |  |  |  |  |  |  |
| ACCESSION AQ089903  |     |  |     |  |  |  |  |  |  |
| VERSION AQ089903.1 GI:3458814   |     |  |     |  |  |  |  |  |  |
| KEYWORDS GSS.   |     |  |     |  |  |  |  |  |  |
| SOURCE human.   |     |  |     |  |  |  |  |  |  |
| ORGANISM Homo sapiens   |     |  |     |  |  |  |  |  |  |
| REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |     |  |     |  |  |  |  |  |  |
| AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |     |  |     |  |  |  |  |  |  |
| 1 (bases 1 to 318)  |     |  |     |  |  |  |  |  |  |
| Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and  |     |  |     |  |  |  |  |  |  |
| Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and   |     |  |     |  |  |  |  |  |  |
| Hood,L.   |     |  |     |  |  |  |  |  |  |
| Sequence-tagged connectors: A sequence approach to mapping and  |     |  |     |  |  |  |  |  |  |
| scanning the human genome   |     |  |     |  |  |  |  |  |  |
| Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)   |     |  |     |  |  |  |  |  |  |
| 99380589  |     |  |     |  |  |  |  |  |  |
| Contact: Mahairas GG, Wallace JC, Hood L  |     |  |     |  |  |  |  |  |  |
| High Throughput Sequencing Center   |     |  |     |  |  |  |  |  |  |
| University of Washington  |     |  |     |  |  |  |  |  |  |
| 401 Queen Anne Avenue North, Seattle, WA 98109, USA   |     |  |     |  |  |  |  |  |  |
| Tel: (206) 616-3618   |     |  |     |  |  |  |  |  |  |
| Fax: (206) 616-3887   |     |  |     |  |  |  |  |  |  |
| Email: jwallace@u.washington.edu  |     |  |     |  |  |  |  |  |  |
| Sequence Tagged Connector   |     |  |     |  |  |  |  |  |  |
| Plate: 3001 row: 0 column: 8  |     |  |     |  |  |  |  |  |  |
| Class: BAC ends   |     |  |     |  |  |  |  |  |  |
| High quality sequence stop: 318.  |     |  |     |  |  |  |  |  |  |
| Location/Qualifiers   |     |  |     |  |  |  |  |  |  |
| 1. 318  |     |  |     |  |  |  |  |  |  |
| /organism="Homo sapiens"  |     |  |     |  |  |  |  |  |  |
| /db_xref="taxon:9606"   |     |  |     |  |  |  |  |  |  |
| /clone="Plate=3001 Col=8 Row=O"   |     |  |     |  |  |  |  |  |  |
| /clone_lib="CIT Approved Human Genomic Sperm Library D"   |     |  |     |  |  |  |  |  |  |
| /sex="male"   |     |  |     |  |  |  |  |  |  |
| /note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in  |     |  |     |  |  |  |  |  |  |
| E-Coli DH10B"   |     |  |     |  |  |  |  |  |  |
| 77 a 66 c 66 g 108 t 1 others   |     |  |     |  |  |  |  |  |  |
| BASE COUNT<br>ORIGIN  |     |  |     |  |  |  |  |  |  |
| Query Match 12.7%; Score 35.6; DB 224; Length 318;<br>Best Local Similarity 54.6%; Pred. NO. 3.6;   |     |  |     |  |  |  |  |  |  |



/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

| Query Match  | 12.5% | Score 35.2    | DB 233        | Length 848 |
|--|-------|---------------|---------------|------------|
| Best Local Similarity  | 55.8% | Pred. No. 5.6 |               |            |
| Matches  | 67    | Conservative  | Mismatches 53 | Indels 0   |
| Gaps   |       |               |               |            |
| 122 gttaaacacagggcattcagttcggacaataacgaccccttagttatcttgggttcgac    | 181   |               |               |            |
| 334 GTTCTGGCCAGGGCAATCAGGTAGGAGAAATAAACGGTATTCAATTAGGAAAAG         | 393   |               |               |            |
| 182 catgtgcattactcgtatttgcgaatgctcctgcctgattatttgaataatcagcaggagac | 241   |               |               |            |
| 204 catgtgcattactcgtatttgcgaatgctcctgcctgattatttgaataatcagcaggagac | 453   |               |               |            |

| RESULT     | 8  |
|------------|--|
| AO418525   | AO418525 487 bp DNA  |
| LOCUS      | RPCI-11-20318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-20318, 23-MAR-1999 |
| DEFINITION | RPCI-11-20318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-20318, 23-MAR-1999 |
|            | DNA sequence.  |

ACCESSION AQ418525  
VERSION AQ418525.1 GI:4476249  
KEYWORDS GSS.  
SOURCE human.

| ORGANISM  | Homo sapiens  |
|-----------|---|
| REFERENCE | 1 (bases 1 to 487)  |
|           | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.       |
|           | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |

|           |   |
|-----------|---|
| REFERENCE | 1 (bases 1 to 487)  |
| AUTHORS   | Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C. |
| TITLE     | Use of BAC End Sequences from Library RpCI-11 for Sequence-Ready            |

Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208

Research Genet cs (info@resgen.com). BAC end search page.  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search.html).  
 Seq primer: T7  
 Class: BAC ends.

```

BASE COUNT      165 a   102 c   96 q   124 t
                /clone_lib="RPCI-11"
                /sex="Male"
                /cell_type="Lymphocytes"
                /note="Vector: pBACE3.6; Site_1: EcoRI;
                RPCI11 Human Male BAC Library"
                Site_2: EcoRI;

```

```

BASE COUNT      165 a   102 c   96 g   124 t
ORIGIN
      12.3%;   Score 34.6;   DB 228;   Length 487;
Query Match
Best Local Similarity 54.3%;   Pred: No. 7.6;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps
QY  108 cccgttgtaaatgctgttagaagacaggcgcattccagttcggacaaatacgaaccccttagtt 167

```

```

Db 127 CTCTGTAGTAATTCGGTCGTCAGGCAATCAGCAGGAGAAATAAAGGCTGTTTC 186
      ||| ||||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| |||
QY 168 attctgggttcaccatgtgcatgtactgtattgttgccaatgtcctgctgattattga 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 AATTAGGAAAAGGAAGTCAAAATTTGTCCTGTTTGCAGATGACATGATTGTTATCTAC 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 228 aatcagca 236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 AAAACCCCA 255
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
LOCUS AQ417598 555 bp DNA GSS 23-MAR-1999
DEFINITION RPCI-11-203C8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203C8,
DNA sequence.
ACCESSION AQ417598
VERSION AQ417598.1 GI:4475441
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 555)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-203C8.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..555
                     /organism="Homo sapiens"
                     /db_xref="GDB:7577623"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-203C8"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT          192 a 114 c 110 g 139 t
ORIGIN
Query Match          12.3%; Score 34.6; DB 228; Length 555;
Best Local Similarity 54.3%; Pred. No. 7.8;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 108 ccctgttagaattgtttaagaacaggcattcagttcgcacaataacgaccccttagtt 167
      ||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 150 CTCTGGTAGAATTCGGTCGTCAGGCAATCAGCAGGAGAAATAAAGGCTGTTTC 209
      ||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||
QY 168 attctgggttcaccatgtgcatgtactgtattgttgccaatgtcctgctgattattga 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 AATTAGGAAAAGGAAGTCAAAATTTGTCCTGTTTGCAGATGACATGATTGTTATCTAC 269
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 228 aatcagca 236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 AAAACCCCA 278
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
LOCUS CNS016H0 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACNI6B20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106734.1 GI:5623558
VERSION AL106734.1
KEYWORDS fruit fly.
SOURCE Plasmid Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelobAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACNI6B20"
                     /note="end : T7"
BASE COUNT          260 a 123 c 159 g 217 t 342 others
ORIGIN
Query Match          12.3%; Score 34.6; DB 219; Length 1101;
Best Local Similarity 30.8%; Pred. No. 8.9;
Matches 57; Conservative 47; Mismatches 81; Indels 0; Gaps 0;

QY 90 tagagtatttaattatgccctgttagaattgtttaagaacaggcattcagttcggaca 149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 TAAATWADTWAWATWBWTAATTTGTTTNRASAAAARTTWGGGAKADWTTTATNTKRGR 725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 150 aatacgaaccccttagttatcttggttcgaccatgtgcactgtactgtatttgccaatg 209
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 726 DGGGGCAACAGYKTADTGATGKAKRSGAARGGWKGTGTRTATATWGGDRKAGW 785
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 210 tctgcctgattatttgaaataacagcaggagcagataaaaactgtggaattttaaagta 269
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 TTGTTKTTRTKAWCKTDKAKTGARTGAGGGAGWBGARAVTGWGATWTDABWBTSDWRGAG 845
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 270 tttgt 274
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 NRTAT 850
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
LOCUS AQ010299 443 bp DNA GSS 21-MAY-1998
DEFINITION HS_2172_B1_MR_G08 ClT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2172 Col=15 Row=N, DNA sequence.
ACCESSION AQ010299
VERSION AQ010299.1 GI:3143369

```

KEYWORDS SOURCE ORGANISM

REFERENCE AUTHORS

TITLE

JOURNAL MEDLINE COMMENT

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2172 row: N column: 15

Class: BAC ends

High quality sequence stop: 443.

Location/Qualifiers

1. 443

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-2172 Col-15 Row-N"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 172 a 91 c 72 g 107 t 1 others

ORIGIN

Query Match 12.1%; Score 34; DB 223; Length 443;

Best Local Similarity 55.7%; Pred. No. 11;

Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 122 gtttaagaacaggcatttcgtagcgaacaaatcgaaccccttagttatcttgggttcgac 181

DB 318 GTTTGGCCAGGCGAGTCAGGCAAGAGAAATAAAGCGTATTCAATAGGAAGAGAA 377

QY 182 catgtgcattactctgtatttgcgaatgcctgcctgatttgcgaatcagca 236

DB 378 GAAGTCAAAATGCCCATGTTTCAGATGACATGATGTTTATNTAGAAACCCCA 432

RESULT 12

LOCUS AQ807176/c

DEFINITION HS\_3249\_A2\_H10\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3249 Col-20 Row-O, DNA sequence.

ACCESSION AQ807176

VERSION AQ807176.1 GI:5724508

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 410)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 3249 row: O column: 20

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 410.

Location/Qualifiers

1. 410

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-3249 Col=20 Row=O"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 93 a 72 c 80 g 164 t 1 others

ORIGIN

Query Match 12.0%; Score 33.8; DB 233; Length 410;

Best Local Similarity 53.4%; Pred. No. 13;

Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 122 gtttaagaacaggcatttcgtagcgaacaaatcgaaccccttagttatcttgggttcgac 181

DB 302 GTTCTGACCCAGGCAATCAGGCAAGAGAAATAAATGAAGATATTCAAATAGGAAGAGAG 243

QY 182 catgtgcattactctgtatttgcgaatgcctgcctgatttgcgaatcagcaggggac 241

DB 242 GAAGTCAAAATATATTTGTTGCAGATGACATGATGGATATTAGAAAACCCCATGTC 183

QY 242 agaataaaactgt 254

DB 182 TCACCCCAAAATCT 170

RESULT 13

LOCUS AG019820

DEFINITION Homo sapiens genomic DNA, 21q region, clone: B2289H10 SpN072(RP), genomic survey sequence.

ACCESSION AG019820

VERSION AG019820.1 GI:6045764

KEYWORDS GSS.

SOURCE Homo sapiens DNA, clone:B2289H10 SpN072(RP).

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 604)

AUTHORS Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA, chromosome 21q

JOURNAL Published Only in DataBase (1999) In press

REFERENCE 2 (bases 1 to 604)

AUTHORS Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (16-OCT-1999) to the DDBJ/EMBL/GenBank databases.

Mahaira Hattori, RIKEN Genomic Sciences Center(GSC) c/o Kitasato University; 1-15-1 Kitasato, Sagamihara 228-8555, Japan (E-mail:hattori@gsc.ims.u-tokyo.ac.jp, Tel:042-778-9923, Fax:042-778-9924)

Location/Qualifiers

1. 604

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="21"

/clone="B2289H10 SpN072(RP)"

/map="21q"



```

BASE COUNT      218 a   98 c   138 g   150 t
ORIGIN

Query Match      12.0%; Score 33.8; DB 219; Length 604;
Best Local Similarity 62.4%; Pred. No. 14;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 50 ccaaacgaaatttaactcgcagtgaccaggggttaatttagagatttataaattatgcc 109
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 CTAAACAAATGAAGTCACCTCTTAACACAGGTGAATTTATAGTGTGCAAAATTATA 370
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 110 ctgttagaattgtttaagaacagg 134
    || || || || || || || || || || || || || || || || || || || ||
Db 371 CTAAATAAAACTGTTTAAAAAAAAGG 395
    || || || || || || || || || || || || || || || || || || || ||

RESULT 14
A0488982      772 bp   DNA      GSS      24-APR-1999
LOCUS      RPCI-11-267P13.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-267P13
DEFINITION      , DNA sequence.
ACCESSION      A0488982
VERSION      A0488982.1 GI:4674856
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 772)
AUTHORS      Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
                ,J.C.
TITLE      Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
                Map Building
JOURNAL      Unpublished (1997)
COMMENT      Other_GSSs: RPCI-11-267P13.TV
                Contact: Shaying Zhao, William Nierman, Mark Adams
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: hbe@tigr.org
                Clones are derived from the human BAC library RPCI-11. For BAC
                library availability, please contact Pieter de Jong
                (pieter@dejong.med.buffalo.edu). Clones may be purchased from
                BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
                Research Genet cs (info@resgen.com). BAC end search page:
                http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
                Seq primer: SP6
                Class: BAC ends.
FEATURES      source
                Location/Qualifiers
                1..772
                /organism="Homo sapiens"
                /db_xref="GB:7602516"
                /db_xref="taxon:9606"
                /clone="RPCI-11-267P13"
                /clone_lib="RPCI-11"
                /sex="Male"
                /cell_type="Lymphocytes"
                /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                RPCI11 Human Male BAC Library"
BASE COUNT      300 a   172 c   122 g   178 t
ORIGIN

Query Match      12.0%; Score 33.8; DB 229; Length 772;
Best Local Similarity 53.4%; Pred. No. 14;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 122 gtttaagacagggcattcagtcgcgacaaatcacgaaccccttagttatcttggttcgac 181
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 514 GTTCTGTCAGGCATCAGGCAGAGAGAAATAACGGATATTTCAGTTAGGAAAGAG 573

BASE COUNT      218 a   98 c   138 g   150 t
ORIGIN

QY 182 catgctgcatgactcgtatttgcgaatgctgctgatttattgaaataacagcaggac 241
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 574 GAAGTGAATTTGTCTCTGTTTGCAGATGACATGACTGTATATTAGAAAACTCCATCGTC 633
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
QY 242 agaataaaactgt 254
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 634 TCAGCCCAATCT 646
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||

RESULT 15
CNS032CO      894 bp   DNA      GSS      18-MAY-2000
LOCUS      Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION      070019 of library G from Tetraodon nigroviridis, genomic survey
                sequence.
ACCESSION      AL267441
VERSION      AL267441.1 GI:7989249
KEYWORDS      GSS: genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM      Tetraodon nigroviridis
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
                Tetraodontidae; Tetraodon.
REFERENCE      1 (bases 1 to 894)
AUTHORS      Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
                Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
                Weissenbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
                freshwater pufferfish Tetraodon nigroviridis
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 894)
AUTHORS      Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
                Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
                Saurin,W. and Weissenbach,J.
TITLE      Human gene number estimate provided by genome wide analysis using
                Tetraodon nigroviridis DNA sequence
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 894)
AUTHORS      Genoscope.
TITLE      Direct Submission
JOURNAL      Submitted (12-APR-2000) to the EMBL/GenBank/DDAJ databases
COMMENT      This sequence is a single read and was generated as part of a large
                scale clone-end sequencing project of the Tetraodon nigroviridis
                genome. For more information, please take a look at
                http://www.genoscope.cns.fr/Tetraodon.
FEATURES      source
                Location/Qualifiers
                1..894
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="070019"
                /clone_lib="G"
                /note="Genoscope sequence ID : CORG070AH10LP1-end : T7"
BASE COUNT      249 a   190 c   181 g   260 t   14 others
ORIGIN

Query Match      12.0%; Score 33.8; DB 221; Length 894;
Best Local Similarity 60.9%; Pred. No. 15;
Matches 53; Conservative 1; Mismatches 33; Indels 0; Gaps 0;

QY 69 tcgattgacccaggtgtaatttagagatttataaattagccctgttaaatgtgttaaga 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 793 TCGAAGCACACGTGTTTACTTTCTTAATAATATACATGCTTGGAGTAGGTGTGCAGA 852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||

QY 129 acaggcattcagttccggacaaatcac 155
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 853 ACAGAGCGTKATGTCGGATGTATAAG 879
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||

Search completed: October 24, 2001, 13:13:48

```

Thu Oct 25 13:08:30 2001

us-09-462-955-1\_copy\_711\_991.rst

Page 10

Job time: 11813 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:27 ; Search time 180.6 Seconds  
(without alignments)  
294.554 Million cell updates/sec

Title: US-09-462-955-1\_COPY\_711\_991  
Perfect score: 281  
Sequence: 1 cacatgtgtggaacagaa.....ttaaagtatgtcatctaa 281

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 70.8  | 25.2        | 1091   | 2  | US-08-418-071-5   |
| 2          | 70.8  | 25.2        | 1096   | 2  | US-08-418-071-4   |
| 3          | 70.8  | 25.2        | 1106   | 2  | US-08-418-071-3   |
| 4          | 65.2  | 23.2        | 1017   | 4  | US-08-793-634B-6  |
| 5          | 61.8  | 22.0        | 1022   | 4  | US-08-793-634B-2  |
| 6          | 44.6  | 15.9        | 1105   | 1  | US-08-202-186-18  |
| 7          | 44.4  | 15.8        | 1103   | 1  | US-08-202-186-16  |
| 8          | 43    | 15.3        | 1110   | 1  | US-08-202-186-14  |
| 9          | 43    | 15.3        | 1111   | 1  | US-08-202-186-12  |
| 10         | 42.8  | 15.2        | 1104   | 1  | US-08-202-186-17  |
| 11         | 41.4  | 14.7        | 982    | 3  | US-08-973-068-28  |
| 12         | 41.4  | 14.7        | 1109   | 1  | US-08-202-186-13  |
| 13         | 41.4  | 14.7        | 1111   | 1  | US-08-202-186-9   |
| 14         | 41.4  | 14.7        | 1111   | 1  | US-08-202-186-24  |
| 15         | 40    | 14.2        | 1111   | 1  | US-08-202-186-10  |
| 16         | 39.8  | 14.2        | 1110   | 1  | US-08-202-186-11  |
| 17         | 39.8  | 14.2        | 1111   | 1  | US-08-202-186-15  |
| 18         | 29.4  | 10.5        | 8920   | 2  | US-08-446-855A-1  |
| 19         | 29.4  | 10.5        | 8920   | 4  | US-09-150-741-1   |
| 20         | 28.8  | 10.2        | 10607  | 1  | US-08-078-090-3   |
| 21         | 27.8  | 9.9         | 2570   | 2  | US-09-056-075-2   |
| 22         | 27.6  | 9.8         | 2899   | 2  | US-08-624-581-2   |
| 23         | 27.6  | 9.8         | 2917   | 2  | US-08-624-581-3   |
| 24         | 27.6  | 9.8         | 2959   | 2  | US-08-624-581-1   |
| 25         | 27.6  | 9.8         | 6476   | 4  | US-09-127-670-5   |
| 26         | 27.4  | 9.8         | 840    | 5  | PCT-US91-08177-12 |
| 27         | 27.4  | 9.8         | 2256   | 6  | 5220013-1         |

|      |      |     |       |   |                   |
|------|------|-----|-------|---|-------------------|
| c 28 | 27.4 | 9.8 | 7323  | 5 | PCT-US91-08177-1  |
| c 29 | 27.2 | 9.7 | 5892  | 3 | US-08-755-587-27  |
| c 30 | 27.2 | 9.7 | 7240  | 3 | US-08-755-587-15  |
| c 31 | 27.2 | 9.7 | 11283 | 2 | US-08-603-753D-3  |
| c 32 | 27.2 | 9.7 | 11283 | 3 | US-09-099-753-3   |
| c 33 | 27.2 | 9.7 | 11283 | 4 | US-08-986-106-3   |
| c 34 | 27.2 | 9.7 | 11385 | 2 | US-08-639-501-1   |
| c 35 | 27.2 | 9.7 | 11385 | 3 | US-09-044-946-1   |
| c 36 | 27.2 | 9.7 | 11385 | 3 | US-09-044-908-1   |
| c 37 | 27   | 9.6 | 1815  | 4 | US-09-586-935-2   |
| c 38 | 27   | 9.6 | 2042  | 2 | US-08-933-821-16  |
| c 39 | 27   | 9.6 | 2042  | 3 | US-08-934-494-5   |
| c 40 | 27   | 9.6 | 2042  | 3 | US-08-960-507-16  |
| c 41 | 27   | 9.6 | 2042  | 3 | US-09-143-068-5   |
| c 42 | 26.8 | 9.5 | 1320  | 2 | US-08-853-659A-12 |
| c 43 | 26.8 | 9.5 | 15512 | 2 | US-08-853-659A-5  |
| c 44 | 26.8 | 9.5 | 15512 | 2 | US-08-853-659A-8  |
| c 45 | 26.8 | 9.5 | 15512 | 2 | US-08-853-659A-63 |

ALIGNMENTS

RESULT 1  
US-08-418-071-5  
; Sequence 5, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh  
; APPLICANT: You, Li-Ru  
; APPLICANT: Soong, Tai-Seng  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/418,071  
; FILING DATE: 06-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: DCB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1091 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; MOLECULE TYPE: subgenomic DNA  
; DESCRIPTION: /desc "BBTV DNA II (clone 2)"  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Banana Bunchy Top Virus (BBTV)  
US-08-418-071-5

Query Match 25.2%; Score 70.8; DB 2; Length 1091;  
 Best Local Similarity 56.2%; Pred. No. 2.5e-14;  
 Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

QY 8 ggtggaacagaaagcgctattgtaccagtagtacatcgagagaccacaaacgaaatttaac 67  
 DB 663 GGTGGAAGACGCTGGATATGATGACATCATACATGATGATGATGATGATGATGATGAT 722  
 QY 68 ctgcatgtaccaggtgtaatttagagtagtatttaattatccctgtttagaattgtttaag 127  
 DB 723 ATTGATATCCCGAAGCAITTCAGATTATCTGAATTATGCGTTATAGAACAAATTAAG 782  
 QY 128 aacaggcattcagtcggacaaatagaacc---ccttagttattcgttgggtcgaccat 184  
 DB 783 AATAGAGTTTAAATAAATACAAAATACGAACCATGTGTGATGAGAAAGATGGACAAAT 842  
 QY 185 gtgcattgtactcgtatttgcgaatgtccctgctgctgattatttgaataatcagcaggacaga 244  
 DB 843 GTCCATGTAATGTTATGCGAAATGTGTGCTGATTTAAATTTTTCAGAAAGATAGA 902  
 QY 245 ataaactgtggaatatttaaaagtattgtgtcatc 278  
 DB 903 ATAAAAATAAATAATGTTGAGAAAGGAAACTTC 936

RESULT 2  
 US-08-418-071-4  
 ; Sequence 4, Application US/08418071  
 ; Patent No. 5846705  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wu, Rey-Yuh  
 ; APPLICANT: You, Li-Ru  
 ; APPLICANT: Soong, Tai-Seng  
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Neave  
 ; STREET: 1251 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: United States  
 ; ZIP: 10020  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 5.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/418,071  
 ; FILING DATE: 06-APR-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haley Jr., James F.  
 ; REGISTRATION NUMBER: 27,794  
 ; REFERENCE/DOCKET NUMBER: DCB-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 596-9000  
 ; TELEFAX: (212) 596-9090  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1096 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: circular  
 ; MOLECULE TYPE: genomic DNA  
 ; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)  
 ; US-08-418-071-4

Query Match 25.2%; Score 70.8; DB 2; Length 1096;  
 Best Local Similarity 56.2%; Pred. No. 2.5e-14;  
 Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

QY 8 ggtggaacagaaagcgctattgtaccagtagtacatcgagagaccacaaacgaaatttaac 67  
 DB 668 GGTGGAAGACGCTGGATATGATGACATCATACATGATGATGATGATGATGATGATGAT 727  
 QY 68 ctgcatgtaccaggtgtaatttagagtagtatttaattatccctgtttagaattgtttaag 127  
 DB 728 ATTGATATCCCGAAGCAITTCAGATTATCTGAATTATGCGTTATAGAACAAATTAAG 787  
 QY 128 aacaggcattcagtcggacaaatagaacc---ccttagttattcgttgggtcgaccat 184  
 DB 788 AATAGAGTTTAAATAAATACAAAATACGAACCATGTGTGATGAGAAAGATGGACAAAT 847  
 QY 185 gtgcattgtactcgtatttgcgaatgtccctgctgctgattatttgaataatcagcaggacaga 244  
 DB 848 GTCCATGTAATGTTATGCGAAATGTGTGCTGATTTAAATTTTTCAGAAAGATAGA 907  
 QY 245 ataaactgtggaatatttaaaagtattgtgtcatc 278  
 DB 908 ATAAAAATAAATAATGTTGAGAAAGGAAACTTC 941

RESULT 3  
 US-08-418-071-3  
 ; Sequence 3, Application US/08418071  
 ; Patent No. 5846705  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wu, Rey-Yuh  
 ; APPLICANT: You, Li-Ru  
 ; APPLICANT: Soong, Tai-Seng  
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Neave  
 ; STREET: 1251 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: United States  
 ; ZIP: 10020  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 5.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/418,071  
 ; FILING DATE: 06-APR-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haley Jr., James F.  
 ; REGISTRATION NUMBER: 27,794  
 ; REFERENCE/DOCKET NUMBER: DCB-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 596-9000  
 ; TELEFAX: (212) 596-9090  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1106 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: circular  
 ; MOLECULE TYPE: genomic DNA  
 ; DESCRIPTION: desc="BBTV DNA I (clone 7-4-2)"  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)  
 ; ORIGINAL SOURCE:  
 ; US-08-418-071-3

```
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-3

Query Match      25.2%; Score 70.8; DB 2; Length 1106;
Best Local Similarity 55.2%; Pred. No. 2.5e-14;
Matches 138; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 20 aaggacgtattgtaccagtcacatcgagagaccccaaaagaaatttaattcctcgatgtacc 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 AATACTTATTCTCTCTATGTGACGACGAGGATCTGACAAAGCATATAGTATTTGATTTCC 736
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 80 aggtgaatttagagatttaaaattgcccctgttagaagtgtttaagaaacagggcattc 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 COTTGTAATCAGGATTTATTAATATATCATGTAAATAGGACATTAAGGATAGGGTTATA 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 140 agttcgacaataacgaaccccttagttatttgcctgttagaagtgtttaagaaacagggcattc 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 GAGACTACTAAATCAACACCCATTAAGATGTTGAATAGGTAATAAATACATGTAAATCGTC 856
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 200 ttgccaatgtcctgtgattttgaaaaatcagcagggagacagaaataaaactgtggaat 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 ATGGCGAATTTTCATGCTGACTTCTGTAATAATCTCCGAAGATGAATAAATAATCAITTTAT 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 260 atttaagta 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 TGCTGAAGAA 926
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-08-793-634B-6
; Sequence 6, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.
; APPLICANT: Larkin, Philip J.
; APPLICANT: Taylor, William C.
; APPLICANT: Marchall, Jerry S.
; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,634B
; FILING DATE: June 9, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-793-634B-6

Query Match      23.2%; Score 65.2; DB 4; Length 1017;
Best Local Similarity 55.2%; Pred. No. 1.8e-12;
Matches 149; Conservative 0; Mismatches 118; Indels 3; Gaps 1;

QY 7 tgggtgaaccagaaagagctattgtaccagtcacatcgagagaccccaaaacgaatttaatt 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 TGGAGGAAGAACACACAGATATGATGCATCTTTGTGACTGTGACCTTAAGAAATAATTGGGT 703
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 67 cctcgatgtaccaggtgtaatttagagattttaaaattatgcctgttagaattgtattaa 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 704 ATTTCATATACCCAGAGTTAGTTTCAGAGTATGTGAATTTATGTTGTAATAGAACAGGTTAA 763
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 127 gaacagggcattcagttcggacaaaatacgaaccccttagttattcctgggttcgacca--- 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 764 GAATAGGGTAATGGTGAATTAAGTATGAGGCATGTGTAATGCGGGATGATAAATCATCC 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 184 tgtgcattactcgtatttggccaatgtcctgctgattttgaaaaatcagcagggacag 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 824 TGTTCATGTAAATTTGTTGTCAAATGTACTCCAGATTTGGGAAAAATTAAGTGAAGATAG 883
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 244 aataaaactgtggaatatttaagtgatgtg 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 884 AATAAAATTAATTCGTTGTTGAAACTCTG 913
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-08-793-634B-2
; Sequence 2, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.
; APPLICANT: Larkin, Philip J.
; APPLICANT: Taylor, William C.
; APPLICANT: Marchall, Jerry S.
; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,634B
; FILING DATE: June 9, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 base pairs
; TYPE: nucleic acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-793-634B-2

Query Match      22.0%; Score 61.8; DB 4; Length 1022;
Best Local Similarity 52.1%; Pred. No. 2.4e-11;
Matches 138; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 2 acatgtgtggaacaggaagacgtatgtaccagtlacatcgaggaccacaaacgaat 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 661 ACAGCCGAGGAGACCCAGGAGTATTATATATATATATATATATATATATATAT 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 ttaatcctcgatgtaccagtgtaatttagagattattaaattatgcccctgttagaattgt 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 721 ATTGGCTTTGATGTCCAGGCTGTTCTTCGGAGAGATGATGAACACTATCAGCGGATGGAGATG 780
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 122 gttaaagaacaggcattcagttcggacaataacgaaccccttagttctgggttcgac 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 781 TTGAAGAACAGAGTTTTTGCAGAGTACAAATATATAGGCCTGTAGATCTTTGTATTAGGAG 840
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 catgtcgtactgtactgttcccaatgtcctgcctgattttaaataatcagcagggac 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 841 TTAGTTCAATTTAATTTGTTTGGCAACGTGGCACCCTGACCCCGCATAGTCAAGTCAAGGAC 900
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 agaataaaactgtggaattatttaa 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 901 AGACTTGTAAATATCAATTTGTGAA 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-18

Query Match      15.9%; Score 44.6; DB 1; Length 1105;
Best Local Similarity 55.4%; Pred. No. 1.2e-05;
Matches 108; Conservative 0; Mismatches 84; Indels 3; Gaps 1;

QY 58 aaatttaatcctcgatgtaccagtgtaatttagagattattaaattatgcccctgttaga 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 760 AATAGTTATATTGATATATCCAGATGCAAGAGGAATATTTAAACTATGTTTATTAGA 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 118 atgtgttaagaacagggcattcagttcggacaataacgaaccccttagttatcttgggtt 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 820 AGAATTTAAGAATGGAATTTATTCAAAGCGGGAATATGAACCCGTTTTCGAAATG---T 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 ccaccatgtcgtactcgttatttgcgaatgtcctgcctgattttaaataatcagcag 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 877 AGAATATGTGGAAGTCATTTGTAATGCTAACTCTCTCCGAAGGAAGGAATCTTTTCAGA 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 ggacagaataaaact 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 937 AGATCGAATAAAGCT 951
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

Query Match      15.8%; Score 44.4; DB 1; Length 1103;
Best Local Similarity 55.8%; Pred. No. 1.4e-05;
Matches 106; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 63 taatcctcgatgtaccagtgtaatttagagattattaaattatgcccctgttagaattgt 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 763 TTATATTTGATATTTCCAGATGCAAGAGGAATATTTAAACTATGTTTATTAGGAAT 822
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

QY 123 ttaagaacaggcattcagttcggacaaatacgaaccccttagttatcttgggttgacc 182  
||||| ||| |  
Db 823 TTAATAATGGAATATTCAAGCGGAAATATGAACCGTTTGAATAATG---TAGAAT 879  
||||| ||| |  
QY 183 atgtcatgtactcgtatttgcgaatctcctgcctgattatttgaataatcagcaggaca 242  
||||| ||| |  
Db 880 ATGTGGAAGTCATTGTAATGCTAACTTCCTTCGGAAGGAATCTTTCTGAAGATC 939  
||||| ||| |  
QY 243 gaataaaact 252  
||||| ||| |  
Db 940 GAATAAAGCT 949  
||||| ||| |

## RESULT 8

US-08-202-186-14  
; Sequence 14, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JEFFERY, Donald D.  
; REGISTRATION NUMBER: 19,980  
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202 672 5300  
; TELEFAX: 202 672 5399  
; TELEX: 904136

INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1110 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
US-08-202-186-14

Query Match 15.3%; Score 43; DB 1; Length 1110;  
Best Local Similarity 54.9%; Pred. No. 4.2e-05;  
Matches 107; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

QY 59 aatttaacccctcagtgaccagggtgaatttagagatttaataatgacctgtagaa 118  
||||| ||| |  
Db 758 ATGTGTTATTTGATATCCCTAGATGCAAGAGGATTTTAAATTTGGTTATTAGAG 817  
||||| ||| |  
QY 119 tgtgtaagaacaggcattcagttcggacaaatacgaaccccttagttatcttgggttc 178  
||||| ||| |  
Db 818 GAATTTAAGAATGGAATAATCAAGCGGAAATATGAACCGTTTGAAGATG---TA 874  
||||| ||| |  
QY 179 gaccatgtcactcgtactcgtatttgcgaatctcctgcctgattatttgaataatcagcagg 238  
||||| ||| |  
Db 875 GAATATGTCGAAGTCATTGTAATGGCTAACTTCCTTCGGAAGGAATCTTTTCTGAA 934  
||||| ||| |  
QY 239 gacagaataaaactg 253  
||||| ||| |

Db 935 GATCGAATAAAGTTG 949  
||||| ||| |

## RESULT 9

US-08-202-186-12  
; Sequence 12, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BURNS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JEFFERY, Donald D.  
; REGISTRATION NUMBER: 19,980  
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202 672 5300  
; TELEFAX: 202 672 5399  
; TELEX: 904136

INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1111 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
US-08-202-186-12

Query Match 15.3%; Score 43; DB 1; Length 1111;  
Best Local Similarity 54.9%; Pred. No. 4.2e-05;  
Matches 107; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

QY 59 aatttaacccctcagtgaccagggtgaatttagagatttaataatgacctgtagaa 118  
||||| ||| |  
Db 759 ATGTGTTATTTGATATTTCAAGATGCAAGAGGATTTTAAATTTGGTTATTAGAA 818  
||||| ||| |  
QY 119 tgtgtaagaacaggcattcagttcggacaaatacgaaccccttagttatcttgggttc 178  
||||| ||| |  
Db 819 GAATTTAAGAATGGAATAATCAAGCGGAAATATGAACCGTTTGAAGATAG---TA 875  
||||| ||| |  
QY 179 gaccatgtcactcgtactcgtatttgcgaatctcctgcctgattatttgaataatcagcagg 238  
||||| ||| |  
Db 876 GAATATGTCGAAGTCATTGTAATGGCTAACTTCCTTCGGAAGGAATCTTTTCTGAA 935  
||||| ||| |  
QY 239 gacagaataaaactg 253  
||||| ||| |  
Db 936 GATCGAATAAAGTTG 950  
||||| ||| |

## RESULT 10

US-08-202-186-17  
; Sequence 17, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:

```

; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-17

Query Match 15.2%; Score 42.8; DB 1; Length 1104;
Best Local Similarity 55.3%; Pred. No. 4.8e-05;
Matches 105; Conservative 0; Mismatches 82; Indels 3; Gaps 1;

QY 63 taatctcgatgtaccaggtgaattagagattttaaattatgcccctgttagaatgt 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 764 TTATATTTCATATCCAGATCAAAGAGGAATATTAAACTATGCTTTATTAGAAGAT 823

QY 123 ttaagacagggcattcagttcgacgaataacgaaccccttagttatcttgggtcgacc 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 824 TTAATAATGGAATTATTCAAGCGGGAATATGAACCCGTTTGAATAATTG---TAGAAT 880

QY 183 atgtcgatgtactcgatttgccaatgtcctgctgatttgaataacagcagggaca 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 881 ATGTGAAGTCATGTGATGGCTACTTCCCTCCGAAGGAATCTTTCTGAAGATC 940

QY 243 gaataaaact 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 941 GAATAAAGCT 950

RESULT 11
US-08-973-068-28
; Sequence 28, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068

```

```

; CURRENT FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; EARLIER FILING DATE: 1996-05-31
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 982
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; US-08-973-068-28

Query Match 14.7%; Score 41.4; DB 3; Length 982;
Best Local Similarity 54.4%; Pred. No. 0.00013;
Matches 106; Conservative 0; Mismatches 86; Indels 3; Gaps 1;

QY 59 aatttaactctcgatgtaccaggtgtaatttagagattttaaattatgcccctgttagaa 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 attgtattatttgatattccaaagatgcaagaggattttaaattatggtttattagag 287

QY 119 tgtgttaagaacagggcatttcgacaaatacgaaccccttagttattcttgggttc 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 gaatttaagaatggaataattcgaacgggaaatatgaacccgttttgaagatag---ta 344

QY 179 gaccatgtcgatgtactcgatttgccaatgtcctgctgatttgaataacagcaggg 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 gaatatgtcgagtcattgttaattgcttaacttcttccgaaggaagaatctttctgaa 404

QY 239 gacagaataaaactg 253
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 gatcgaataaaagtgtg 419

RESULT 12
US-08-202-186-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular

```





Search completed: October 24, 2001, 10:00:29  
Job time: 214 sec

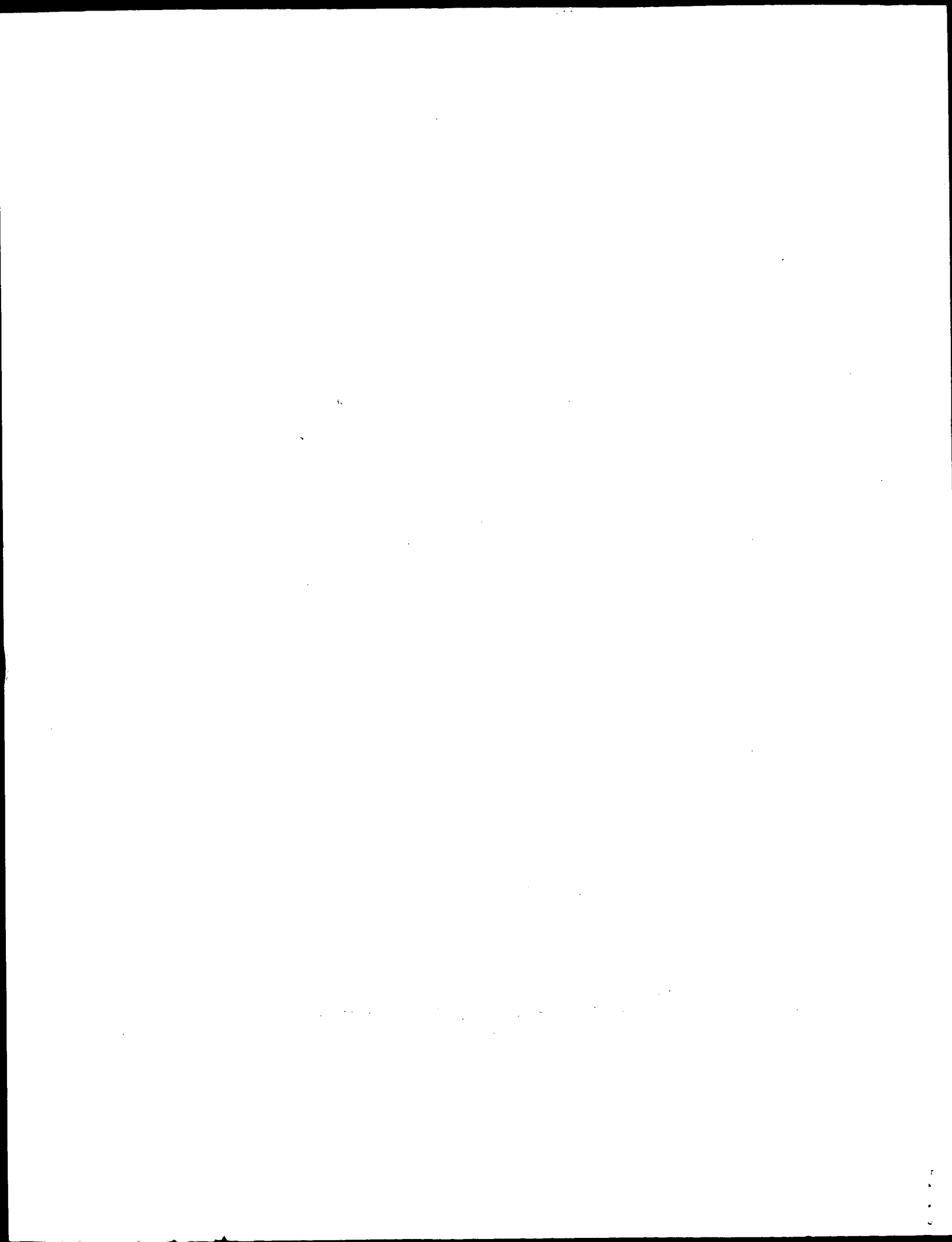
RESULT 15  
US-08-202-186-10  
; Sequence 10, Application US/08202186  
; Patent No. 5756708  
; GENERAL INFORMATION:  
; APPLICANT: KARAN, Mirko  
; APPLICANT: BORNIS, Thomas M.  
; APPLICANT: DALE, James L.  
; APPLICANT: HARDING, Robert M.  
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202,186  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JEFFERY, Donald D.  
; REGISTRATION NUMBER: 19,980  
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202 672 5300  
; TELEFAX: 202 672 5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1111 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; US-08-202-186-10

|     | Query Match           | 14.2%   | Score 40;          | DB 1;          | Length 1111; |            |
|-----|-----------------------|---|--------------------|----------------|--------------|------------|
|     | Best Local Similarity | 53.5%;  | Pred. No. 0.00041; |                |              |            |
|     | Matches 107;          | Conservative  | 0;                 | Mismatches 90; | Indels       | 3; Gaps 1; |
| Qy  | 54                    | aacgaatttaactcctcgatgtaccgcggtgtaatttagtagtatttaaattatgcctgt    | 113                |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
| Ddb | 754                   | aagatattgtttattttgattttccaagatgcacagaggatttttaattatgggttat      | 813                |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
| Qy  | 114                   | tagaatgtgttaagaacagggcattcagttcgacacaaatacgaaccccttagttattcttg  | 173                |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
| Ddb | 814                   | tagaggaaatttttaagagtggaataattcaaacgcgggaatatgaaccggttttgaagatag | 873                |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
| Qy  | 174                   | ggttcgaccattgctactcgtatttgccaatgtcctgcctgattatttgaataatca       | 233                |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
| Ddb | 874                   | ---TAGAATATGTGCAACTCATTTGTAATGCTTACCTTCCTCCGAGGAGGAATCTTTT      | 930                |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
| Qy  | 234                   | gcaggacagaataaaaactg  | 253                |                |              |            |
|     |                       |   |                    |                |              |            |
|     |                       |   |                    |                |              |            |
| Ddb | 931                   | CTGAAGATCGAATAAAGTTG  | 950                |                |              |            |

Thu Oct 25 13:08:29 2001

us-09-462-955-1\_copy\_711\_991.rni

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:25 ; Search time 393.52 Seconds  
(without alignments)  
49.464 Million cell updates/sec

Title: US-09-462-955-2

Perfect score: 31

Sequence: 1 agccgcggggtaatactagccccgcgcgt 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N\_Geneseq\_0601:\*

2: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:\*

3: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:\*

4: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:\*

5: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:\*

6: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:\*

7: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:\*

8: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:\*

9: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:\*

10: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:\*

11: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:\*

12: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:\*

13: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:\*

14: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:\*

15: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:\*

16: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:\*

17: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:\*

18: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:\*

19: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:\*

20: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:\*

21: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:\*

22: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:\*

23: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 31    | 100.0       | 31     | 15 | AAQ56057    |
| 2          | 31    | 100.0       | 31     | 20 | AAQ56057    |
| 3          | 31    | 100.0       | 31     | 20 | AAQ56057    |
| C 4        | 19.8  | 63.9        | 31     | 15 | AAQ56057    |
| C 5        | 19.8  | 63.9        | 31     | 20 | AAQ56057    |
| C 6        | 19.8  | 63.9        | 31     | 20 | AAQ56057    |
| C 7        | 19.8  | 63.9        | 31     | 20 | AAQ56057    |
| C 8        | 18.4  | 59.4        | 415    | 21 | AAQ56057    |
| C 9        | 18.2  | 58.7        | 60     | 18 | AAQ56057    |
| C 10       | 18.2  | 58.7        | 60     | 18 | AAQ56057    |
| C 11       | 18.2  | 58.7        | 352    | 18 | AAQ56057    |

|      |      |      |        |    |          |                    |
|------|------|------|--------|----|----------|--------------------|
| C 12 | 18.2 | 58.7 | 533    | 18 | AAQ56057 | Banana bunchy top  |
| C 13 | 18.2 | 58.7 | 623    | 18 | AAQ56057 | Banana bunchy top  |
| C 14 | 18.2 | 58.7 | 630    | 20 | AAQ56057 | Nucleotide sequenc |
| C 15 | 18.2 | 58.7 | 684    | 21 | AAQ56057 | Human prostate can |
| C 16 | 18.2 | 58.7 | 1075   | 18 | AAQ56057 | Banana bunchy top  |
| C 17 | 18.2 | 58.7 | 1075   | 18 | AAQ56057 | Banana bunchy top  |
| C 18 | 18.2 | 58.7 | 1075   | 18 | AAQ56057 | Banana bunchy top  |
| C 19 | 18.2 | 58.7 | 1089   | 18 | AAQ56057 | Banana bunchy top  |
| C 20 | 18.2 | 58.7 | 1089   | 18 | AAQ56057 | Banana bunchy top  |
| C 21 | 18.2 | 58.7 | 1089   | 18 | AAQ56057 | Banana bunchy top  |
| C 22 | 18.2 | 58.7 | 1110   | 19 | AAQ56057 | Banana bunchy top  |
| C 23 | 17.8 | 57.4 | 389    | 21 | AAQ56057 | Banana bunchy top  |
| C 24 | 17.8 | 57.4 | 811    | 21 | AAQ56057 | N. meningitidis pa |
| C 25 | 17.8 | 57.4 | 1284   | 8  | AAQ56057 | Maize gibberellin  |
| C 26 | 17.8 | 57.4 | 1284   | 11 | AAQ56057 | N-myc 1 proto-onco |
| C 27 | 17.8 | 57.4 | 1284   | 19 | AAQ56057 | N-myc 1 clone. A   |
| C 28 | 17.8 | 57.4 | 2974   | 14 | AAQ56057 | Genomic DNA sequen |
| C 29 | 17.4 | 56.1 | 1438   | 21 | AAQ56057 | Nucleotide sequenc |
| C 30 | 17.2 | 55.5 | 1340   | 21 | AAQ56057 | Wnt-4AF and Wnt-5c |
| C 31 | 17.2 | 55.5 | 2284   | 22 | AAQ56057 | Human MyoD1 gene p |
| C 32 | 17.2 | 55.5 | 2450   | 20 | AAQ56057 | Nucleotide sequenc |
| C 33 | 17.2 | 55.5 | 4086   | 14 | AAQ56057 | Human myoD gene an |
| C 34 | 17.2 | 55.5 | 12019  | 20 | AAQ56057 | Alcaligenes sp. Po |
| C 35 | 17.2 | 55.5 | 437    | 21 | AAQ56057 | Aspergillus oryzae |
| C 36 | 17.2 | 55.5 | 480    | 21 | AAQ56057 | Myrtaceae microsat |
| C 37 | 17.2 | 55.5 | 20387  | 19 | AAQ56057 | HSV-2 strain S85 C |
| C 38 | 17.2 | 55.5 | 26338  | 19 | AAQ56057 | HSV-2 strain S85 C |
| C 39 | 17.2 | 55.5 | 96988  | 21 | AAQ56057 | BAC containing rep |
| C 40 | 17.2 | 55.5 | 117213 | 19 | AAQ56057 | HSV-2 strain S85 C |
| C 41 | 16.8 | 54.2 | 477    | 20 | AAQ56057 | Polynucleotide seq |
| C 42 | 16.8 | 54.2 | 521    | 21 | AAQ56057 | Trichoderma reesei |
| C 43 | 16.8 | 54.2 | 1719   | 19 | AAQ56057 | Human par-4 gene.  |
| C 44 | 16.8 | 54.2 | 4140   | 13 | AAQ56057 | Encodes acid alpha |
| C 45 | 16.8 | 54.2 | 8467   | 20 | AAQ56057 | Polynucleotide seq |

ALIGNMENTS

RESULT 1

AAQ56057

ID AAO56057 standard; DNA; 31 BP.

XX AAO56057;

XX AAO56057;

XX 12-AUG-1994 (first entry)

XX Coconut Foliar Decay Virus promoter stem-loop.

XX Coconut Foliar Decay Virus; CPDV; strong promoter; tissue-specific;

XX phloem-specific; stem-loop structure; transgenic plant; ds.

XX Coconut Foliar Decay Virus.

XX Key Location/Qualifiers

FT stem\_loop 1..31

FT /\*tag= a

FT /function= promoter

FT /note= "loop has homology to geminivirus sequence"

XX DE4306832-C.

XX 24-FEB-1994.

XX 04-MAR-1993; 93DE-4306832.

XX 04-MAR-1993; 93DE-4306832.

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Becker D, Randles JW, Rohde W, Salamini F;

XX WPI; 1994-058406/08.

```

XX Use of coconut foliar decay virus DNA as promoter - for
PT tissue-specific gene expression in transgenic plants
XX
XX Claim 1; Fig 2; 8pp; German.
XX
XX A DNA fragment from the CFV genome can be used as a phloem-specific
CC promoter in the construction of transgenic plants. The promoter is
CC strong; it has 30-50% of the activity of the CaMV 35S promoter in
CC tobacco plants.
XX
XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgcgt 31
    |||||
DB 1 agccgcggggtaatactagcccccgcgcgt 31

RESULT 2
AAx02688
ID AAX02688 standard; DNA; 31 BP.
XX
AC AAX02688;
XX
XX 10-MAY-1999 (first entry)
XX
XX Gemini virus DNA fragment stem loop.
XX
XX Stem loop; coconut foliar decay virus; CFV; bacterial promoter;
KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX
XX Gemini virus.
XX
XX Key Location/Qualifiers
FH stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX
XX DE19730502-A1.
XX
XX 21-JAN-1999.
XX
XX 16-JUL-1997; 97DE-1030502.
XX
XX 16-JUL-1997; 97DE-1030502.
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX WPI; 1999-096863/09.
XX
XX Coconut foliar decay virus promoters - for gene expression in
XX plants, bacteria and yeasts
XX
XX Disclosure; Fig 2; 14pp; German.
XX
XX This invention describes a coconut foliar decay virus (CFV) DNA fragment
XX that includes the stem-loop structure of CFV DNA but lacks the
XX translation start codons of open reading frames ORF1 and/or ORF2. The
XX new CFV DNA fragment is useful as a bacterial or yeast promoter, as a
XX promoter for tissue-specific (especially phloem-specific) gene expression
XX in plants and for production of chimeric constructs for transient or
XX CC
XX
XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgcgt 31
    |||||
DB 1 agccgcggggtaatactagcccccgcgcgt 31

RESULT 3
AAx02686
ID AAX02686 standard; DNA; 31 BP.
XX
AC AAX02686;
XX
XX 10-MAY-1999 (first entry)
XX
XX Gemini virus DNA fragment stem loop.
XX
XX Stem loop; coconut foliar decay virus; CFV; bacterial promoter;
KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX
XX Gemini virus.
XX
XX Key Location/Qualifiers
FH stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX
XX DE19730535-A1.
XX
XX 21-JAN-1999.
XX
XX 16-JUL-1997; 97DE-1030535.
XX
XX 16-JUL-1997; 97DE-1030535.
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX WPI; 1999-096867/09.
XX
XX Coconut foliar decay virus promoters - for gene expression in
XX plants, bacteria and yeasts
XX
XX Disclosure; Fig 2; 12pp; German.
XX
XX This invention describes a coconut foliar decay virus (CFV) DNA fragment
XX that includes the stem-loop structure of CFV DNA but lacks the
XX translation start codons of open reading frames ORF1 and/or ORF2. The
XX new CFV DNA fragment is useful as a bacterial or yeast promoter, as a
XX promoter for tissue-specific (especially phloem-specific) gene expression
XX in plants and for production of chimeric constructs for transient or
XX CC
XX
XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagcccccgcgcgt 31  
 |||||  
 Db 1 agccgcggggtaataactagcccccgcgcgt 31

RESULT 4  
 AAO56057/c  
 ID AAO56057 standard; DNA; 31 BP.  
 XX  
 AC AAO56057;  
 XX  
 DT 12-AUG-1994 (first entry)  
 XX  
 DE Coconut Foliar Decay Virus promoter stem-loop.  
 XX  
 KW Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;  
 KW phloem-specific; stem-loop structure; transgenic plant; ds.  
 XX  
 OS Coconut Foliar Decay Virus.  
 XX  
 FH Key Location/Qualifiers  
 FT stem\_loop 1..31  
 FT /\*tag= a  
 FT /function= promoter  
 FT /note= "loop has homology to geminivirus sequence"

DE4306832-C.  
 XX  
 PN 24-FEB-1994.  
 XX  
 PD 04-MAR-1993; 93DE-4306832.  
 XX  
 PF 04-MAR-1993; 93DE-4306832.  
 XX  
 PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 XX  
 PA Becker D, Randles JW, Rohde W, Salamini F;  
 XX  
 DR WPI; 1994-058406/08.  
 XX

Use of coconut foliar decay virus DNA as promoter - for  
 tissue-specific gene expression in transgenic plants  
 XX  
 Claim 1; Fig 2; 8pp; German.  
 XX

A DNA fragment from the CFDV genome can be used as a phloem-specific  
 promoter in the construction of transgenic plants. The promoter is  
 strong; it has 30-50% of the activity of the CamV 35S promoter in  
 tobacco plants.  
 XX

Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 63.9%; Score 19.8; DB 15; Length 31;  
 Best Local Similarity 77.4%; Pred. No. 3.9;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagcccccgcgcgt 31  
 |||||  
 Db 31 AGCCGCGGGGCTAGTATTACCCCGCGGCT 1

RESULT 5  
 AAX02688/c  
 ID AAX02688 standard; DNA; 31 BP.  
 XX  
 AC AAX02688;  
 XX

10-MAY-1999 (first entry)  
 XX  
 DE Gemini virus DNA fragment stem loop.

XX  
 KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;  
 KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.  
 XX  
 OS Gemini virus.  
 XX

FH Key Location/Qualifiers  
 FT stem\_loop 1..31  
 FT /\*tag= a  
 FT misc\_binding 1..11  
 FT /\*tag= b  
 FT /note= "Region binds to nucleotides 21 to 31"  
 FT misc\_binding 21..31  
 FT /\*tag= c  
 FT /note= "Region binds to nucleotides 1 to 11"

DE19730502-A1.  
 XX  
 PN 21-JAN-1999.  
 XX  
 PD 16-JUL-1997; 97DE-1030502.  
 XX  
 PF 16-JUL-1997; 97DE-1030502.  
 XX  
 PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 XX

PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;  
 XX  
 DR WPI; 1999-096863/09.  
 XX  
 DR Coconut foliar decay virus promoters - for gene expression in  
 bacteria and yeasts  
 XX  
 PS Disclosure; Fig 2; 14pp; German.  
 XX

This invention describes a coconut foliar decay virus (CFDV) DNA fragment  
 that includes the stem-loop structure of CFDV DNA but lacks the  
 translation start codons of open reading frames ORF1 and/or ORF2. The  
 new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a  
 promoter for tissue-specific (especially phloem-specific) gene expression  
 in plants and for production of chimeric constructs for transient or  
 stable expression. Certain fragments of CFDV DNA have stronger promoter  
 activity in E. coli than the CamV 35S promoter.  
 XX

Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 63.9%; Score 19.8; DB 20; Length 31;  
 Best Local Similarity 77.4%; Pred. No. 3.9;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagcccccgcgcgt 31  
 |||||  
 Db 31 AGCCGCGGGGCTAGTATTACCCCGCGGCT 1

RESULT 6  
 AAX02686/c  
 ID AAX02686 standard; DNA; 31 BP.  
 XX  
 AC AAX02686;  
 XX

10-MAY-1999 (first entry)  
 XX  
 DE Gemini virus DNA fragment stem loop.

Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;  
 KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.  
 XX  
 OS Gemini virus.  
 XX

FH Key Location/Qualifiers  
 FT stem\_loop 1..31

us-09-462-955-2.rng

Thu Oct 25 13:08:40 2001

```

FT      misc_binding      /*tag= a.
FT      1..11.
FT      /*tag= b
FT      /*note= "Region binds to nucleotides 21 to 31"
FT      21..31
FT      /*tag= c
FT      /*note= "Region binds to nucleotides 1 to 11"
FT      1..11
PN      DE19730535-A1.
XX      21-JAN-1999.
XX      16-JUL-1997; 97DE-1030535.
XX      16-JUL-1997; 97DE-1030535.
XX      (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX      Becker D, Hehn A, Randles JW, Rohde W, Salami F;
XX      WPI; 1999-096867/09.
XX      Coconut foliar decay virus promoters - for gene expression in
XX      plants, bacteria and yeasts
XX      Disclosure; Fig 2; 12pp; German.
XX      This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX      that includes the stem-loop structure of CFDV DNA but lacks the
XX      translation start codons of open reading frames ORF1 and/or ORF2. The
XX      new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX      promoter for tissue-specific (especially phloem-specific) gene expression
XX      in plants and for production of chimeric constructs for transient or
XX      stable expression. Certain fragments of CFDV DNA have stronger promoter
XX      activity in E. coli than the CAMV 35S promoter.
XX      Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
XX      Query Match 53.9%; Score 19.8; DB 20; Length 31;
XX      Best Local Similarity 77.4%; Pred. No. 3.9;
XX      Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 agccgcggggtaatactagccccgcggct 31
DB      31 AGCCGCGGGGCTAGTATTACCCCGGGCT 1

RESULT 7
AAV27112
ID      AAV27112 standard; DNA; 35099 BP.
XX      AC      AAV27112;
XX      28-SEP-1998 (first entry)
XX      DE      Adenovirus 17.
XX      Adeno virus 17; Ad17; vector; gene therapy; cystic fibrosis;
XX      alpha-antitrypsin deficiency; respiratory disease; ss.
XX      Mastadenovirus 17.
XX      WO9822609-A1.
XX      28-MAY-1998.
XX      20-NOV-1997; 97WO-US21494.
XX      20-NOV-1996; 96US-0752760.
XX      (GENZ ) GENZYME CORP.
XX

```

```

PI      Armentano DE, Gregory RJ, Smith AE;
XX      WPI; 1998-312493/27.
XX      Recombinant adenovirus type 2 vector - useful for targetting
XX      biologically active proteins, used to, e.g. treat cystic fibrosis
XX      Disclosure; Page 28-38; 67pp; English.
XX      This is the complete nucleotide (nt) sequence of adenovirus 17
XX      (Ad17). A claimed chimeric adenoviral vector comprises a nt
XX      sequence of an adenovirus in which all or part of a gene encoding
XX      a protein facilitating attachment to, or internalisation into, a
XX      target mammalian cell is replaced by all or part of the
XX      corresponding gene for a second adenovirus from subgroup D,
XX      preferably selected from Ad9, Ad15, Ad17, Ad19, Ad22, Ad26,
XX      Ad27, Ad28, Ad30 and Ad39. The vector further comprises a
XX      transgene operably linked to a eukaryotic promoter to allow
XX      expression in a mammalian cell. The replaced gene preferably
XX      encodes an Ad fibre such as Ad17 fibre (see AAV27113) and/or an Ad
XX      penton base such as Ad17 penton base (see AAV27114). The chimeric
XX      vector is used to target biologically active proteins to airway
XX      epithelial cells, especially for gene therapy of diseases such as
XX      cystic fibrosis or alpha-antitrypsin deficiency. The vectors are
XX      based on the discovery that proteins from the subgroup D viruses,
XX      a group not normally associated with human respiratory diseases,
XX      can effectively bind and internalise within human airway
XX      epithelial cells.
XX      Sequence 35099 BP; 7988 A; 9877 C; 9978 G; 7252 T; 4 other;
XX      Query Match 60.6%; Score 18.8; DB 19; Length 35099;
XX      Best Local Similarity 76.7%; Pred. No. 22;
XX      Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 agccgcggggtaatactagccccgcggc 30
DB      15104 accgctgggtattactagggccagcagc 15133

RESULT 8
AAV09348
ID      AAV09348 standard; cDNA; 415 BP.
XX      AC      AAV09348;
XX      13-MAR-2001 (first entry)
XX      DE      Fusarium venenatum EST SEQ ID NO:1871.
XX      Multiple gene expression; filamentous fungal cell; EST;
XX      expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX      Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX      culture condition; environmental stress; spore morphogenesis;
XX      metabolic pathway engineering; catabolic pathway engineering; ss.
XX      Fusarium venenatum.
XX      WO2000056762-A2.
XX      28-SEP-2000.
XX      22-MAR-2000; 2000WO-US07781.
XX      22-MAR-1999; 99US-0273623.
XX      (NOVO ) NOVO NORDISK BIOTECH INC.
XX      (NOVO ) NOVO NORDISK AS.
XX      Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX      WPI; 2000-594572/56.
XX

```



XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 PS Claim 86; Page 1082; 3161pp; English.  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from  
 CC *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and  
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are  
 CC all specifically claimed in the present invention.  
 XX  
 SQ Sequence 415 BP; 95 A; 122 C; 113 G; 79 T; 6 other;

Query Match 59.4%; Score 18.4; DB 21; Length 415;  
 Best Local Similarity 78.6%; Pred. No. 21;  
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 2 gccgcgggggtaatactagccccgcgg 29  
 III III III III III III III  
 Db 38 gccacgagagtactacgagccgcgg 65

RESULT 9  
 AAT49396/c  
 ID AAT49396 standard; DNA; 60 BP.  
 XX  
 AC AAT49396;  
 XX  
 DT 23-AUG-1997 (first entry)  
 XX  
 DE Banana bunchy top virus DNA component 6 stem-loop common region.  
 XX  
 KW BBTv; intergenic region; promoter; transgenic plant; ss.  
 XX  
 OS Banana bunchy top virus.  
 XX  
 FH Key Location/Qualifiers  
 FT stem\_loop 17..47  
 FT /\*tag= a  
 FT misc\_binding 17..26  
 FT /\*tag= b  
 FT /\*note= "binds to nts 47..56"  
 FT misc\_structure 27..37  
 FT /\*tag= c  
 FT /\*note= "loop"  
 FT misc\_binding 38..47  
 FT /\*tag= d  
 FT /\*note= "binds to nts 26..35"  
 XX  
 PN WO9638554-A1.

PD 05-DEC-1996.  
 XX  
 PF 31-MAY-1996; 96WO-AU00335.  
 XX  
 PR 31-MAY-1995; 95AU-0003285.  
 XX  
 PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.  
 XX  
 PI Becker DK, Beetham PR, Dale JL, Dugdale B, Hafner GJ;  
 PI Harding RM;  
 XX  
 DR WPI; 1997-034368/03.  
 XX  
 PT DNA from intergenic region of banana bunchy top virus DNA component  
 PT - useful for promoting, enhancing, regulating or modifying  
 PT transcription of a non-BBTv gene in transgenic plants  
 XX  
 PS Claim 4; Fig 3; 80pp; English.  
 XX  
 CC Stem-loop common regions (AAT49391-96) are conserved in DNA  
 CC components 1-6 (see also AAT49386-90) of banana bunchy top virus  
 CC virus (BBTV). Each component has an 11-nucleotide loop sequence  
 CC of which 9 consecutive nucleotides are conserved between all 6  
 CC components, and a 10 bp stem sequence of which 14 nucleotides are  
 CC conserved. Intergenic regions (see also AAT49399-409) of components  
 CC 1-6, including regions contg. these stem-loop common regions, are  
 CC useful for promoting, enhancing, regulating or modifying  
 CC transcription of non-BBTv genes in monocotyledon or dicotyledon  
 CC transgenic plants.  
 XX  
 SQ Sequence 60 BP; 13 A; 17 C; 18 G; 12 T; 0 other;

Query Match 58.7%; Score 18.2; DB 18; Length 60;  
 Best Local Similarity 74.2%; Pred. No. 22;  
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 agccgcgggggtaatactagccccgcggct 31  
 III III III III III III III  
 Db 47 AGCAGCGGGGGTAATAATAAGTCCCGGTGCT 17

RESULT 10  
 AAT49393/c  
 ID AAT49393 standard; DNA; 69 BP.  
 XX  
 AC AAT49393;  
 XX  
 DT 23-AUG-1997 (first entry)  
 XX  
 DE Banana bunchy top virus DNA component 3 stem-loop common region.  
 XX  
 KW BBTv; intergenic region; promoter; transgenic plant; ss.  
 XX  
 OS Banana bunchy top virus.  
 XX  
 FH Key Location/Qualifiers  
 FT stem\_loop 26..56  
 FT /\*tag= a  
 FT misc\_binding 26..35  
 FT /\*tag= b  
 FT /\*note= "binds to nts 47..56"  
 FT misc\_structure 36..46  
 FT /\*tag= c  
 FT /\*note= "loop"  
 FT misc\_binding 47..56  
 FT /\*tag= d  
 FT /\*note= "binds to nts 26..35"  
 XX  
 PN WO9638554-A1.  
 XX  
 PD 05-DEC-1996.  
 XX



components 1-6 (see also AAT9386-90) of banana bunchy top virus (BBTV). Intergenic region bbtvpro6 (AAT9402) comprises a BBTV component 6 insert in pBT6.1, a plasmid obtd. by PCR amplification (see also AAT9416-17), subcloning as a Pst-BamHI fragment into pUC19 and further cloning as a HindIII-BamHI fragment into pBI101.3. The component 6 intergenic region has promoter activity comparable to that of the 800 bp CMV 35S promoter from pBI121. CC BBTV intergenic regions are useful for promoting, enhancing, CC regulating or modifying transcription of non-BBTV genes in CC monocotyledon or dicotyledon transgenic plants. The non-BBTV gene CC is e.g. an insecticide resistance gene, herbicide resistance gene, CC or growth promoting gene.

Sequence 623 BP; 223 A; 101 C; 137 G; 162 T; 0 other;

Query Match 58.7%; Score 18.2; DB 18; Length 623;  
Best Local Similarity 74.2%; Pred. No. 27;  
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps

QY 1 agccgcggggttaactagaccgcccgctt 31  
||| ||||||||| ||| ||||| |||  
Db 374 AGCAGCGGGGTAAATATACTCCCCGTC 344

RESULT 14  
AAX34684/C  
ID AAX34684 standard; DNA; 630 BP.  
XX  
AC AAX34684;  
XX  
DT 02-JUL-1999 (first entry)  
XX  
DE Nucleotide sequence of BBTV DNA-6 intergenic region.  
XX  
KW Promoter; Banana Bunchy Top Virus; BBTV; gene transcription; ubil;  
KW polyubiquitin l; banana; ss.  
XX  
OS Banana bunchy top virus.  
XX  
PN WO9915646-A1.  
XX  
PD 01-APR-1999.  
XX  
PF 21-SEP-1998; 98WO-AU00786.  
XX  
PR 30-JUN-1998; 98AU-0004423.  
PR 19-SEP-1997; 97AU-0009339.  
XX  
PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.  
XX  
PI Becker DK, Dale JL, Dugdale B, Harding RM, Hermann SR;  
XX WPI; 1999-254706/21.  
DR  
PT Banana Bunchy Top Virus promoter constructs  
XX  
PS Examples; Fig 1; 84pp; English.  
XX  
CC The invention relates to DNA promoter sequences derived from components  
CC of Banana Bunchy Top Virus (BBTV). The DNA molecule includes a promoter  
CC sequence derived from an untranslated portion of any one BBTV components  
CC and is adaptable for promoting transcription of a cloned gene in a plant  
CC cell. The promoter, included in a DNA chimeric vector, is useful for the  
CC expression of a gene in a plant cell. The inclusion of the polyubiquitin  
CC 1 (ubil) intron sequence into BBTV promoter constructs enhances promoter  
CC activity. The present sequence represents the nucleotide sequence of BBTV  
CC DNA-6 intergenic region.  
XX  
SQ Sequence 630 BP; 227 A; 101 C; 138 G; 164 T; 0 other;

Best Local Similarity 74.2%; Pred. No. 27;  
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 203 AGCCGGGGGAGAACTACAGTCCCGCGGCT 173

Search completed: October 24, 2001, 10:07:26  
Job time: 631 sec

OY 1 agccgggggtaataactagcccccgcgct 31  
||| ||||| ||| ||| ||| |||  
DB 378 AGCAGCGGGGTAAATAGTCCCGCGTCT 348

RESULT 15  
AAFI6045/C  
ID AAFI6045 standard; cDNA; 684 BP.  
XX AC AAFI6045;  
XX DT 13-MAR-2001 (first entry)  
XX DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:480.  
XX KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;  
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease; ss.  
XX OS Homo sapiens.

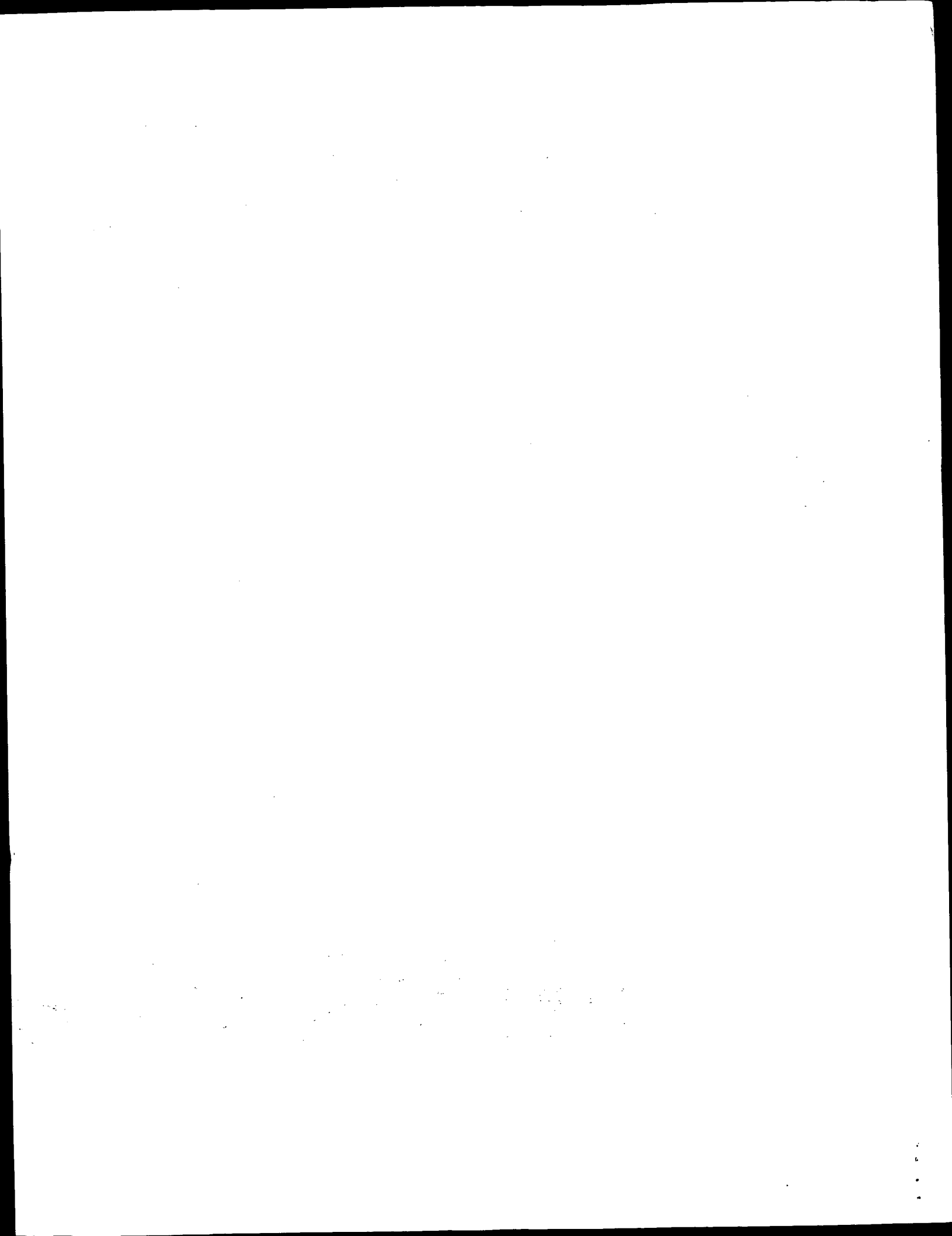
XX WO200055174-A1.  
XX PD 21-SEP-2000.  
XX PF 08-MAR-2000; 2000WO-US05988.  
XX PR 12-MAR-1999; 99US-0124270.  
XX {HUMA-} HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX PI Rosen CA, Ruben SM;  
XX WPI; 2000-587513/55.  
XX P-PSDB; AAB56842.  
XX PT Prostate cancer associated gene sequences, referred to as prostate  
XX cancer antigens, useful for treatment, prevention, and diagnosis of  
XX disorders such as prostate cancer -  
XX Claim 1; Page 952; 2338pp; English.

XX AAFI5566 to AAFI6505 encode the human prostate cancer associated  
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
XX The prostate cancer antigens can have neuroprotective, cytostatic,  
XX cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
XX nephrotropic, antiinfective, gynaecological and antibacterial activities,  
XX and can be used in gene therapy. The prostate cancer antigen  
XX polynucleotides may be used for detection of prostate cancer, chromosome  
XX identification, as chromosome markers, and for numerous other diagnostic  
XX or research purposes. The prostate cancer antigens may be used to treat  
XX disorders such as neural, immune, muscular, reproductive,  
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
XX disorders, wounds, and infectious diseases. AAFI6506 to AAFI6514 to  
XX AAB57303 represent sequences used in the exemplification of the present  
XX invention.  
XX SQ Sequence 684 BP; 111 A; 208 C; 248 G; 112 T; 5 other;

Query Match 58.7%; Score 18.2; DB 21; Length 684;  
Best Local Similarity 74.2%; Pred. No. 28;  
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 agccgggggtaataactagcccccgcgct 31  
||| ||||| ||| ||| ||| |||





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:48 ; Search time 5479.82 Seconds  
(without alignments)  
53.476 Million cell updates/sec

Title: US-09-462-955-2

Perfect score: 31

Sequence: 1 agccgcgggggtaataactagccccgcggct 31

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estomi:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estrol:\*  
83: em\_estrol2:\*  
84: em\_estrol3:\*  
85: em\_estrol4:\*  
86: em\_estrol5:\*  
87: em\_estrol6:\*  
88: em\_estrol7:\*  
89: em\_estrol8:\*  
90: em\_estrol9:\*  
91: em\_estrol10:\*  
92: em\_estrol11:\*  
93: em\_estrol12:\*  
94: em\_estrol13:\*  
95: em\_estrol14:\*  
96: em\_estrol15:\*  
97: em\_estrol16:\*  
98: em\_estrol17:\*  
99: em\_estrol18:\*  
100: em\_estrol19:\*  
101: em\_estrol20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

117: gb\_est48:\*  
 118: gb\_est49:\*  
 119: gb\_est50:\*  
 120: gb\_est51:\*  
 121: gb\_est52:\*  
 122: gb\_est53:\*  
 123: gb\_est54:\*  
 124: gb\_est55:\*  
 125: gb\_est56:\*  
 126: gb\_est57:\*  
 127: gb\_est58:\*  
 128: gb\_est59:\*  
 129: gb\_est60:\*  
 130: gb\_est61:\*  
 131: gb\_est62:\*  
 132: gb\_est63:\*  
 133: gb\_est64:\*  
 134: gb\_est65:\*  
 135: gb\_est66:\*  
 136: gb\_est67:\*  
 137: gb\_est68:\*  
 138: gb\_est69:\*  
 139: gb\_est70:\*  
 140: gb\_est71:\*  
 141: gb\_est72:\*  
 142: gb\_est73:\*  
 143: gb\_est74:\*  
 144: gb\_est75:\*  
 145: gb\_est76:\*  
 146: gb\_est77:\*  
 147: gb\_est78:\*  
 148: gb\_est79:\*  
 149: gb\_est80:\*  
 150: gb\_est81:\*  
 151: gb\_est82:\*  
 152: gb\_est83:\*  
 153: gb\_est84:\*  
 154: gb\_est85:\*  
 155: gb\_est86:\*  
 156: gb\_est87:\*  
 157: gb\_est88:\*  
 158: gb\_est89:\*  
 159: gb\_est90:\*  
 160: gb\_est91:\*  
 161: gb\_est92:\*  
 162: gb\_est93:\*  
 163: gb\_est94:\*  
 164: gb\_est95:\*  
 165: gb\_est96:\*  
 166: gb\_est97:\*  
 167: gb\_est98:\*  
 168: gb\_est99:\*  
 169: gb\_est100:\*  
 170: gb\_est101:\*  
 171: gb\_est102:\*  
 172: gb\_est103:\*  
 173: gb\_est104:\*  
 174: gb\_est105:\*  
 175: gb\_est106:\*  
 176: gb\_est107:\*  
 177: gb\_est108:\*  
 178: gb\_est109:\*  
 179: gb\_est110:\*  
 180: gb\_est111:\*  
 181: gb\_est112:\*  
 182: gb\_est113:\*  
 183: gb\_est114:\*  
 184: gb\_est115:\*  
 185: gb\_est116:\*  
 186: gb\_est117:\*  
 187: gb\_est118:\*  
 188: gb\_est119:\*  
 189: gb\_est120:\*

190: gb\_est121:\*  
 191: gb\_est122:\*  
 192: gb\_est123:\*  
 193: gb\_est124:\*  
 194: gb\_est125:\*  
 195: gb\_est126:\*  
 196: gb\_est127:\*  
 197: gb\_est128:\*  
 198: gb\_est129:\*  
 199: gb\_est130:\*  
 200: gb\_est131:\*  
 201: gb\_est132:\*  
 202: gb\_est133:\*  
 203: gb\_est134:\*  
 204: gb\_est135:\*  
 205: gb\_est136:\*  
 206: gb\_est137:\*  
 207: gb\_est138:\*  
 208: gb\_est139:\*  
 209: gb\_est140:\*  
 210: gb\_est141:\*  
 211: gb\_est142:\*  
 212: gb\_est143:\*  
 213: gb\_est144:\*  
 214: gb\_est145:\*  
 215: gb\_est146:\*  
 216: gb\_est147:\*  
 217: gb\_est148:\*  
 218: gb\_est149:\*  
 219: gb\_est150:\*  
 220: gb\_est151:\*  
 221: gb\_est152:\*  
 222: gb\_est153:\*  
 223: gb\_est154:\*  
 224: gb\_est155:\*  
 225: gb\_est156:\*  
 226: gb\_est157:\*  
 227: gb\_est158:\*  
 228: gb\_est159:\*  
 229: gb\_est160:\*  
 230: gb\_est161:\*  
 231: gb\_est162:\*  
 232: gb\_est163:\*  
 233: gb\_est164:\*  
 234: gb\_est165:\*  
 235: gb\_est166:\*  
 236: gb\_est167:\*  
 237: gb\_est168:\*  
 238: gb\_est169:\*  
 239: gb\_est170:\*  
 240: gb\_est171:\*  
 241: gb\_est172:\*  
 242: gb\_est173:\*  
 243: gb\_est174:\*  
 244: gb\_est175:\*  
 245: gb\_est176:\*  
 246: gb\_est177:\*  
 247: gb\_est178:\*  
 248: gb\_est179:\*  
 249: gb\_est180:\*  
 250: gb\_est181:\*  
 251: gb\_est182:\*  
 252: gb\_est183:\*  
 253: gb\_est184:\*  
 254: gb\_est185:\*  
 255: gb\_est186:\*  
 256: gb\_est187:\*  
 257: gb\_est188:\*  
 258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

| Result No. | Score | Query Match | Length | DB  | ID       | Description        |
|------------|-------|-------------|--------|-----|----------|--------------------|
| 1          | 21.4  | 69.0        | 905    | 220 | CNS03052 | AL221807 Tetraodon |
| 2          | 21.4  | 69.0        | 1072   | 220 | CNS02GRK | AL196697 Tetraodon |
| 3          | 20.4  | 65.8        | 876    | 142 | BE968958 | BE968958 601649921 |
| 4          | 20.4  | 64.5        | 1843   | 144 | BF128533 | BF128533 601810627 |
| 5          | 19.8  | 63.9        | 865    | 141 | BE902577 | BE902577 601677225 |
| 6          | 19.8  | 63.9        | 890    | 168 | BF959233 | BF959233 602080459 |
| 7          | 19.6  | 63.2        | 1287   | 120 | AW731193 | AW731193 CA_Ea001  |
| 8          | 19.4  | 62.6        | 629    | 32  | AV722569 | AV722569 AV722569  |
| 9          | 19.4  | 62.6        | 714    | 106 | AU006051 | AU006051 AU006051  |
| 10         | 19.4  | 62.6        | 797    | 153 | BG401207 | BG401207 602465296 |
| 11         | 19.4  | 62.6        | 836    | 144 | BF126760 | BF126760 601650551 |
| 12         | 19.4  | 62.6        | 897    | 146 | BF300720 | BF300720 602031903 |
| 13         | 19.4  | 62.6        | 983    | 221 | CNS04KIP | AL294874 Tetraodon |
| 14         | 19.4  | 61.3        | 627    | 165 | BE275297 | BE275297 601122139 |
| 15         | 19.4  | 61.3        | 705    | 143 | BF032195 | BF032195 601452609 |
| 16         | 19.4  | 61.3        | 715    | 235 | AQ938642 | AQ938642 NLI-C08R  |
| 17         | 19.4  | 61.3        | 734    | 175 | BG281907 | BG281907 602402672 |
| 18         | 19.4  | 61.3        | 824    | 168 | BF696175 | BF696175 602124688 |
| 19         | 19.4  | 61.3        | 902    | 175 | BG281858 | BG281858 602403088 |
| 20         | 19.4  | 61.3        | 958    | 172 | BF982728 | BF982728 602304933 |
| 21         | 19.4  | 61.3        | 1160   | 172 | BF971171 | BF971171 602270763 |
| 22         | 18.8  | 60.6        | 501    | 155 | BG561147 | BG561147 EtESted79 |
| 23         | 18.8  | 60.6        | 537    | 19  | AI352913 | AI352913 MB73-4B P |
| 24         | 18.8  | 60.6        | 909    | 142 | BE959324 | BE959324 601654233 |
| 25         | 18.8  | 60.6        | 912    | 146 | BF263427 | BF263427 HV_Cea000 |
| 26         | 18.8  | 60.6        | 944    | 140 | BE779793 | BE779793 601465480 |
| 27         | 18.8  | 60.6        | 971    | 222 | CNS04Y22 | AL312452 Tetraodon |
| 28         | 18.8  | 60.6        | 1030   | 137 | BF562027 | BF562027 601345668 |
| 29         | 18.8  | 60.6        | 1072   | 172 | BG030208 | BG030208 602297502 |
| 30         | 18.6  | 60.0        | 431    | 155 | BG557433 | BG557433 EMI_44_A1 |
| 31         | 18.6  | 60.0        | 760    | 251 | AZ900170 | AZ900170 RPT-24-1  |
| 32         | 18.6  | 60.0        | 930    | 105 | AL520966 | AL520966 AL520966  |
| 33         | 18.6  | 60.0        | 954    | 138 | BE620637 | BE620637 601483609 |
| 34         | 18.4  | 59.4        | 115    | 166 | BE348526 | BE348526 ht71e09.x |
| 35         | 18.4  | 59.4        | 224    | 104 | A1969538 | A1969538 wz68c09.x |
| 36         | 18.4  | 59.4        | 277    | 161 | BE577683 | BE577683 BB577683  |
| 37         | 18.4  | 59.4        | 284    | 117 | AW548262 | AW548262 L0033G11- |
| 38         | 18.4  | 59.4        | 319    | 113 | AW204439 | AW204439 UI-H-B11- |
| 39         | 18.4  | 59.4        | 321    | 143 | BF058466 | BF058466 7k31a07.x |
| 40         | 18.4  | 59.4        | 340    | 229 | AQ473306 | AQ473306 C1PBI-El- |
| 41         | 18.4  | 59.4        | 361    | 108 | AU171662 | AU171662 AU171662  |
| 42         | 18.4  | 59.4        | 369    | 21  | A1523972 | A1523972 tg98h02.x |
| 43         | 18.4  | 59.4        | 401    | 108 | AU170664 | AU170664 AU170664  |
| 44         | 18.4  | 59.4        | 447    | 234 | AQ853114 | AQ853114 LMAJFV1_1 |
| 45         | 18.4  | 59.4        | 466    | 138 | BE676163 | BE676163 7f24b10.x |

## ALIGNMENTS

|            |  |
|------------|--|
| RESULT 1   |  |
| CNS03052   |  |
| LOCUS      | 905 bp DNA GSS 15-MAY-2000   |
| DEFINITION | Tetraodon nigroviridis genome survey sequence T7 end of clone 183N12 of library G from Tetraodon nigroviridis, genomic survey sequence.  |
| ACCESSION  | AL221807   |
| VERSION    | AL221807.1 GI:7880626  |
| KEYWORDS   | GSS; genome survey sequence.   |
| SOURCE     | Tetraodon nigroviridis.  |
| ORGANISM   | Tetraodon nigroviridis.<br>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Tetraodontidae; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodon. |
| REFERENCE  | 1 (bases 1 to 905)   |
| AUTHORS    | Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,   |

|   |   |
|---|---|
| Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.<br>Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  |   |
| Unpublished   |   |
| 2 (bases 1 to 905)  |   |
| Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.<br>Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence |   |
| Unpublished   |   |
| 3 (bases 1 to 905)  |   |
| Direct Submission   |   |
| Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases   |   |
| This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.   |   |
| Location/Qualifiers   |   |
| 1..905  |   |
| /organism="Tetraodon nigroviridis"  |   |
| /db_xref="taxon:99883"  |   |
| /clone="183N12"   |   |
| /note="Genoscope sequence ID : C0AG183DG06LPI-end : T7"   |   |
| BASE COUNT  | 138 a 299 c 282 g 177 t 9 Others  |
| ORIGIN  |   |
| Query Match   | 69.0%; Score 21.4; DB 220; Length 905;  |
| Best Local Similarity   | 80.6%; Pred. No. 24;  |
| Matches   | 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;   |
| QY  | 1 agccgcggggtaactagcccgccgct 31<br>   |
| Db  | 482 AGCCGCGAGGCTACTACCGCGCCCGCCGCT 512  |
| RESULT 2  |   |
| CNS02GRK  |   |
| LOCUS   | 1072 bp DNA GSS 13-MAY-2000   |
| DEFINITION  | Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 138009 of library G from Tetraodon nigroviridis, genomic survey sequence.  |
| ACCESSION   | AL196697  |
| VERSION   | AL196697.1 GI:7834847   |
| KEYWORDS  | GSS; genome survey sequence.  |
| SOURCE  | Tetraodon nigroviridis.   |
| ORGANISM  | Tetraodon nigroviridis.<br>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Tetraodontidae; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodon.                                |
| REFERENCE   | 1 (bases 1 to 1072)   |
| AUTHORS   | Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.<br>Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis |
| Unpublished   |   |
| 2 (bases 1 to 1072)   |   |
| Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.<br>Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence |   |
| Unpublished   |   |
| 3 (bases 1 to 1072)   |   |
| Direct Submission   |   |
| Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases   |   |



BE902577  
 VERSION BE902577.1 GI:10392908  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 865)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: L1CM841 row: b column: 08  
 High quality sequence stop: 675.  
 FEATURES  
 source Location/Qualifiers  
 1..865  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3959719"  
 /clone\_lib="NIH\_MGC\_21"  
 /tissue\_type="choriocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 204 a 225 c 287 g 149 t  
 ORIGIN  
 Query Match 63.9%; Score 19.8; DB 141; Length 865;  
 Best Local Similarity 77.4%; Pred. No. 1.2e-02;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 agccgcggggtaatactagccccgcgcgt 31  
 ||||| ||||| ||||| ||||| |||||  
 Db 727 AGCCCGGGGTTATATCGGGCCCTGCCT 757  
 RESULT 6  
 LOCUS BF695233 890 bp mRNA EST 22-DEC-2000  
 DEFINITION 602080459f1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:4244751 5', mRNA sequence.  
 ACCESSION BF695233  
 VERSION BF695233.1 GI:11980641  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 890)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM1057 row: f column: 16  
 High quality sequence stop: 334.  
 FEATURES  
 source Location/Qualifiers  
 1..890  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4244751"  
 /clone\_lib="NIH\_MGC\_81"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCCATATGGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGGCGGCGGCACATG-dt(30)EN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."  
 BASE COUNT 308 a 253 c 155 g 174 t  
 ORIGIN

Query Match 63.9%; Score 19.8; DB 168; Length 890;  
 Best Local Similarity 77.4%; Pred. No. 1.2e-02;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 agccgcggggtaatactagccccgcgcgt 31  
 ||||| ||||| ||||| ||||| |||||  
 Db 685 AGCCACGGGCTAATAAGGCGCGCGCGCT 715  
 RESULT 7  
 LOCUS AW731193 1287 bp mRNA EST 08-MAR-2001  
 DEFINITION GA\_Ea0010K01 Gossypium arboreum 7-10 dpa fiber library Gossypium  
 ACCESSION AW731193  
 VERSION AW731193.1 GI:7628851  
 KEYWORDS EST.  
 SOURCE Gossypium arboreum.  
 ORGANISM Gossypium arboreum.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
 REFERENCE 1 (bases 1 to 1287)  
 AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
 TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)  
 High quality sequence stop: 1287.  
 FEATURES  
 source Location/Qualifiers  
 1..1287  
 /organism="Gossypium arboreum"  
 /strain="AKA"  
 /cultivar="8400"  
 /db\_xref="taxon:29729"  
 /clone="GA\_Ea0010K01"  
 /clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
 /tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
 /lab\_host="E. coli"

```

/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      234 a 382 c 450 g 214 t 7 others
ORIGIN

Query Match      63.2%; Score 19.6; DB 120; Length 1287;
Best Local Similarity 84.6%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 gcgggggtaatactagcccgccg 30
    ||||| ||||| ||||| ||||| |||||
Db 406 GCGGGGTAGTACTCGCCGCCGCCGC 431

RESULT 8
AV722569 629 bp mRNA EST 16-OCT-2000
LOCUS AV722569 HTB Homo sapiens cDNA clone HTBANE01 5', mRNA sequence.
DEFINITION AV722569
ACCESSION AV722569.1 GI:10825189
VERSION AV722569.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 629)
AUTHORS Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z., and Han, Z.
TITLE Homo sapiens cDNA HTB clones
JOURNAL Unpublished (2000)
CONTACT Zeng, Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..629
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTBANE01"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      185 a 124 c 129 g 190 t 1 others
ORIGIN

Query Match      62.6%; Score 19.4; DB 32; Length 629;
Best Local Similarity 79.3%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 gccgggggtaatactagcccgccg 30
    ||||| ||||| ||||| ||||| |||||
Db 597 GCGGTGCGTAATAGCAGCCGCCGTGC 625

RESULT 9
AU006051/c 714 bp mRNA EST 19-JAN-1999
LOCUS AU006051 Bombyx mori p50(Daizo) Bombyx mori cDNA clone ww40411,
DEFINITION AU006051
ACCESSION AU006051
VERSION AU006051.1 GI:4163435
KEYWORDS EST.
SOURCE domestic silkworm.

```

```

Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
; Bombycoidea; Bombycidae; Bombyx.
1 (bases 1 to 714)
Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
Established of cDNA database of Bombyx mori
Unpublished (1999)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
PROJECT = 'CREST project by JST'.
FEATURES
source
Location/Qualifiers
1..714
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/clone="ww40411"
/clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT      173 a 183 c 191 g 166 t 1 others
ORIGIN

Query Match      62.6%; Score 19.4; DB 106; Length 714;
Best Local Similarity 79.3%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccgccg 29
    ||||| ||||| ||||| ||||| |||||
Db 398 AGCCGCGGTAGTAGTCGCGCTCGCGG 370

RESULT 10
BG401207/c 797 bp mRNA EST 12-MAR-2001
LOCUS BG401207 NTH_MGC_75 Homo sapiens cDNA clone IMAGE:4593646 5',
DEFINITION BG401207
ACCESSION BG401207.1 GI:13294655
VERSION BG401207.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
cDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1333 row: o column: 23
High quality sequence stop: 586.
FEATURES
source
Location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4593646"
/clone_lib="NTH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGCCGACATG-dT(30)BN-3' (where B = A,

```

C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 209 a 205 c 193 g 190 t  
 ORIGIN

Query Match 62.6%; Score 19.4; DB 153; Length 797;  
 Best Local Similarity 79.3%; Pred. No. 1.8e+02;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 ccgcggggtaatactagcccccgcgcct 31  
 ||||| ||||| ||||| ||||| |||||

Db 738 CCGCGGGGAATGCTTGCCCGGCGCT 710

RESULT 11  
 BF126760 836 bp mRNA 24-OCT-2000  
 LOCUS 601605051R1 NIH\_MGC\_76 Homo sapiens cDNA clone IMAGE:3934092 3',  
 DEFINITION mRNA sequence.  
 ACCESSION BF126760  
 VERSION BF126760.1 GI:10965800  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 836)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LLCW774 row: f column: 13  
 High quality sequence start: 8  
 High quality sequence stop: 11.  
 Location/Qualifiers  
 1..836  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:3934092"  
 /clone\_lib="NIH\_MGC\_76"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggccctcgcc); Site\_2: SfiI (ggccattagcc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CAGCGCCATTATGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGGAGCGCGGCATG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.85  
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH\_MGC Library."

FEATURES  
 source

Query Match 62.6%; Score 19.4; DB 144; Length 836;  
 Best Local Similarity 79.3%; Pred. No. 1.8e+02;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

BASE COUNT 172 a 235 c 275 g 151 t 3 others  
 ORIGIN

QY 2 ccgcggggtaatactagcccccgcgc 30  
 ||||| ||||| ||||| ||||| |||||

Db 353 GCCCGGGGCAAACTAGGCGCGCGGC 381

RESULT 12  
 BF300720

LOCUS 602031903F1 NCI\_CGAP\_9G2 Mus musculus cDNA clone IMAGE:4166784 5',  
 DEFINITION mRNA sequence.

ACCESSION BF300720  
 VERSION BF300720.1 GI:11247243  
 KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 897)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LLAM9456 row: f column: 01  
 High quality sequence stop: 571.  
 Location/Qualifiers

1..897  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_lib="IMAGE:4166784"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site\_1:  
 NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo  
 dt. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 266 a 216 c 234 g 180 t 1 others  
 ORIGIN

FEATURES  
 source

Query Match 62.6%; Score 19.4; DB 146; Length 897;  
 Best Local Similarity 79.3%; Pred. No. 1.8e+02;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ccgcggggtaatactagcccccgcgc 30  
 ||||| ||||| ||||| ||||| |||||

Db 660 GCCCGGGGAACACACCGCGCGGC 688

RESULT 13  
 CNS04KIP

LOCUS 983 bp DNA 21-MAY-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
 11619 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.

ACCESSION AL294874  
 VERSION AL294874.1 GI:8033454  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 REFERENCE 1 (bases 1 to 983)

AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

us-09-462-955-2.rst

```

/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACACG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
115 a 164 c 203 g 145 t

```

|                       |       |                   |        |            |
|-----------------------|-------|-------------------|--------|------------|
| Query Match           | 61.3% | Score 19          | DB 165 | Length 627 |
| Best Local Similarity | 81.5% | Pred. No. 2.7e+02 |        |            |
| Matches               | 22    | Mismatches        | 5      | Indels 0   |
| Conservative          |       |                   |        | Gaps 0     |

|            |   |
|------------|---|
| RESULT     | 15  |
| BF032195   |   |
| LOCUS      | 705 bp mRNA EST 20-OCT-2000                                     |
| DEFINITION | 601452609F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3856394 5' |
| ACCESSION  | BF032195  |
| VERSION    | BF032195.1 GI:10739907  |
| KEYWORDS   | EST.  |
| SOURCE     | human.  |

|           |   |
|-----------|---|
| ORGANISM  | Homo sapiens  |
| REFERENCE | Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Euteleostomi: Mammalia; Eutheria: Primates: Catarrhini: Hominoidea: Homo. |
| AUTHORS   | 1 (bases 1 to 705)  |
| TITLE     | NTH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .   |
| JOURNAL   | National Institutes of Health, Mammalian Gene Collection (MGC)  |
| COMMENT   | Unpublished (1999)<br>Contact: Robert Strausberg, Ph.D.   |

Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM9585 row: a column: 03  
High quality sequence start: 302.

```

/db_xref="taxon:9606"
/clone="IMAGE:3856394"
/clone.lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pCMV-Sport6; Site_1: Nott; Site_2: Sali; Cloned unidirectionally. Primer: Oligo d Average insert size 1.8 kb. Library constructed by Li Technology"

```

| Query Match           | 61.3% | Score 19           | DB 143 | Length 705                      |
|-----------------------|-------|--------------------|--------|---------------------------------|
| Best Local Similarity | 81.5% | Score: No. 2.7e+02 |        |                                 |
| Matches               | 22    | Conservative       | 0      | Mismatches 5; Indels 0; Gaps 0; |

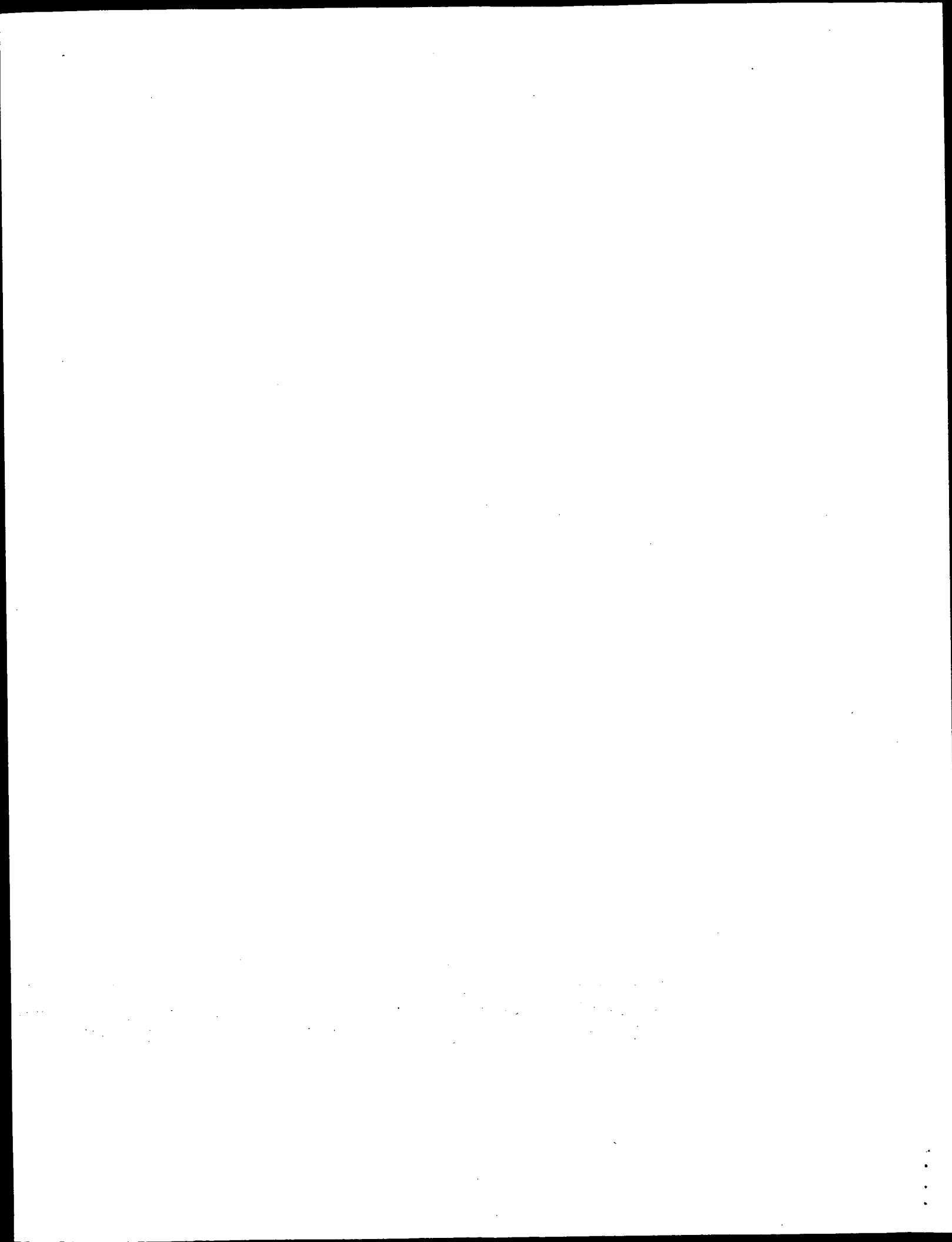
|    |     |                             |     |
|----|-----|-----------------------------|-----|
| Qy | 3   | ccgcgggggttaatactagcccccggg | 29  |
| pb | 679 | ccgcgggggggaaactatcccccggg  | 705 |

Thu Oct 25 13:08:43 2001

us-09-462-955-2.rst

Page 9

Search completed: October 24, 2001, 13:13:52  
Job time: 11817 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw.model

Run on: October 24, 2001, 10:00:29 ; Search time 180.6 seconds  
(without alignments)  
32.495 Million cell updates/sec

Title: US-09-462-955-2

Perfect score: 31

Sequence: 1 agccgcggggtaataactagcccgccgctt 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description      |
|------------|-------|---------|--------------|-------|------------------|
| 1          | 18.8  | 60.6    | 35081        | 2     | US-08-752-760A-1 |
| 2          | 18.2  | 58.7    | 60           | 3     | US-08-973-068-15 |
| 3          | 18.2  | 58.7    | 69           | 3     | US-08-973-068-12 |
| 4          | 18.2  | 58.7    | 352          | 3     | US-08-973-068-30 |
| 5          | 18.2  | 58.7    | 547          | 3     | US-08-973-068-24 |
| 6          | 18.2  | 58.7    | 622          | 3     | US-08-973-068-29 |
| 7          | 18.2  | 58.7    | 624          | 3     | US-08-973-068-27 |
| 8          | 18.2  | 58.7    | 1075         | 1     | US-08-202-186-20 |
| 9          | 18.2  | 58.7    | 1075         | 3     | US-08-973-068-2  |
| 10         | 18.2  | 58.7    | 1089         | 3     | US-08-202-186-23 |
| 11         | 18.2  | 58.7    | 1089         | 3     | US-08-973-068-8  |
| 12         | 18.2  | 58.7    | 1110         | 1     | US-08-202-186-14 |
| 13         | 17.8  | 57.4    | 2974         | 1     | US-08-208-486-2  |
| 14         | 17.2  | 55.5    | 2284         | 4     | US-09-193-792-1  |
| 15         | 17.2  | 55.5    | 4086         | 1     | US-08-313-181-1  |
| 16         | 16.6  | 53.5    | 69           | 3     | US-08-973-068-10 |
| 17         | 16.6  | 53.5    | 69           | 3     | US-08-973-068-13 |
| 18         | 16.6  | 53.5    | 69           | 3     | US-08-973-068-14 |
| 19         | 16.6  | 53.5    | 250          | 3     | US-08-973-068-22 |
| 20         | 16.6  | 53.5    | 532          | 3     | US-08-973-068-26 |
| 21         | 16.6  | 53.5    | 689          | 3     | US-08-973-068-25 |
| 22         | 16.6  | 53.5    | 982          | 3     | US-08-973-068-28 |
| 23         | 16.6  | 53.5    | 1018         | 1     | US-08-202-186-22 |
| 24         | 16.6  | 53.5    | 1018         | 3     | US-08-973-068-6  |
| 25         | 16.6  | 53.5    | 1043         | 1     | US-08-202-186-21 |
| 26         | 16.6  | 53.5    | 1043         | 3     | US-08-973-068-4  |
| 27         | 16.6  | 53.5    | 1103         | 1     | US-08-202-186-16 |

|      |      |      |      |   |                   |                   |
|------|------|------|------|---|-------------------|-------------------|
| c 28 | 16.6 | 53.5 | 1104 | 1 | US-08-202-186-17  | Sequence 17, Appl |
| c 29 | 16.6 | 53.5 | 1105 | 1 | US-08-202-186-18  | Sequence 18, Appl |
| c 30 | 16.6 | 53.5 | 1109 | 1 | US-08-202-186-13  | Sequence 13, Appl |
| c 31 | 16.6 | 53.5 | 1110 | 1 | US-08-202-186-11  | Sequence 11, Appl |
| c 32 | 16.6 | 53.5 | 1111 | 1 | US-08-202-186-9   | Sequence 9, Appl  |
| c 33 | 16.6 | 53.5 | 1111 | 1 | US-08-202-186-10  | Sequence 10, Appl |
| c 34 | 16.6 | 53.5 | 1111 | 1 | US-08-202-186-12  | Sequence 12, Appl |
| c 35 | 16.6 | 53.5 | 1111 | 1 | US-08-202-186-15  | Sequence 15, Appl |
| c 36 | 16.6 | 53.5 | 1111 | 1 | US-08-202-186-24  | Sequence 24, Appl |
| c 37 | 16.6 | 53.5 | 1381 | 2 | US-08-950-449A-13 | Sequence 13, Appl |
| c 38 | 16.6 | 53.5 | 1423 | 1 | US-08-469-421-13  | Sequence 13, Appl |
| c 39 | 16.6 | 53.5 | 1423 | 1 | US-08-250-975-13  | Sequence 13, Appl |
| c 40 | 16.6 | 53.5 | 1423 | 2 | US-08-605-002A-13 | Sequence 13, Appl |
| c 41 | 16.6 | 53.5 | 1423 | 5 | PCT-US94-10529-13 | Sequence 13, Appl |
| c 42 | 16.6 | 53.5 | 3774 | 2 | US-08-950-449A-11 | Sequence 11, Appl |
| c 43 | 16.6 | 53.5 | 3813 | 1 | US-08-469-421-11  | Sequence 11, Appl |
| c 44 | 16.6 | 53.5 | 3813 | 1 | US-08-250-975-11  | Sequence 11, Appl |
| c 45 | 16.6 | 53.5 | 3813 | 2 | US-08-605-002A-11 | Sequence 11, Appl |

ALIGNMENTS

RESULT 1  
US-08-752-760A-1  
; Sequence 1, Application US/08752760A  
; Patent No. 5877011  
; GENERAL INFORMATION:  
; APPLICANT: Armentano, Donna  
; APPLICANT: Gregory, Richard J.  
; APPLICANT: Smith, Alan E.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Baker & Botts, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752.760A  
; FILING DATE: 20-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seide, Rochelle K  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: A31385  
; TELEPHONE: 212-705-5000  
; TELEFAX: 212-705-5020  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 35081 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-752-760A-1

Query Match 60.6% Score 18.8; DB 2; Length 35081;  
Best Local Similarity 76.7%; Pred. No. 17;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

us-09-462-955-2.rni

Thu Oct 25 13:08:42 2001

QY 1 agccgcgggggtaatactagcccccgcgcgc 30  
 DB 15096 ACCGCTGGGGTATTACTAGCCCGCAGC 15125

RESULT 2  
 US-08-973-068-15/c  
 ; Sequence 15, Application US/08973068  
 ; Patent No. 6127604  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dale, James Langham  
 ; APPLICANT: Harding, Robert Maxwell  
 ; APPLICANT: Dugdale, Benjamin  
 ; APPLICANT: Beetham, Peter Ronald  
 ; APPLICANT: Hafner, Gregory John  
 ; APPLICANT: Becker, Douglas Kenneth  
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS  
 ; FILE REFERENCE: 09657/002001  
 ; CURRENT APPLICATION NUMBER: US/08/973,068  
 ; CURRENT FILING DATE: 1998-03-12  
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335  
 ; EARLIER FILING DATE: 1996-05-31  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 15  
 ; LENGTH: 60  
 ; TYPE: DNA  
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)  
 US-08-973-068-15

Query Match 58.7%; Score 18.2; DB 3; Length 60;  
 Best Local Similarity 74.2%; Pred. No. 19;  
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 31  
 DB 47 AGCAGCGGGGGTAAATAAGTATAGTCCCGCGTCT 17

RESULT 3  
 US-08-973-068-12/c  
 ; Sequence 12, Application US/08973068  
 ; Patent No. 6127604  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dale, James Langham  
 ; APPLICANT: Harding, Robert Maxwell  
 ; APPLICANT: Dugdale, Benjamin  
 ; APPLICANT: Beetham, Peter Ronald  
 ; APPLICANT: Hafner, Gregory John  
 ; APPLICANT: Becker, Douglas Kenneth  
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS  
 ; FILE REFERENCE: 09657/002001  
 ; CURRENT APPLICATION NUMBER: US/08/973,068  
 ; CURRENT FILING DATE: 1998-03-12  
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335  
 ; EARLIER FILING DATE: 1996-05-31  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 12  
 ; LENGTH: 69  
 ; TYPE: DNA  
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)  
 US-08-973-068-12

Query Match 58.7%; Score 18.2; DB 3; Length 69;  
 Best Local Similarity 74.2%; Pred. No. 19;  
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 31  
 DB 56 AGCGCTGGGGTAAATAAGTATAGTCCCGCAGC 26

RESULT 4  
 US-08-973-068-30/c  
 ; Sequence 30, Application US/08973068  
 ; Patent No. 6127604  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dale, James Langham  
 ; APPLICANT: Harding, Robert Maxwell  
 ; APPLICANT: Dugdale, Benjamin  
 ; APPLICANT: Beetham, Peter Ronald  
 ; APPLICANT: Hafner, Gregory John  
 ; APPLICANT: Becker, Douglas Kenneth  
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS  
 ; FILE REFERENCE: 09657/002001  
 ; CURRENT APPLICATION NUMBER: US/08/973,068  
 ; CURRENT FILING DATE: 1998-03-12  
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335  
 ; EARLIER FILING DATE: 1996-05-31  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 30  
 ; LENGTH: 352  
 ; TYPE: DNA  
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)  
 US-08-973-068-30

Query Match 58.7%; Score 18.2; DB 3; Length 352;  
 Best Local Similarity 74.2%; Pred. No. 22;  
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 31  
 DB 104 AGCAGCGGGGGTAAATAAGTATAGTCCCGCGTCT 74

RESULT 5  
 US-08-973-068-24/c  
 ; Sequence 24, Application US/08973068  
 ; Patent No. 6127604  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dale, James Langham  
 ; APPLICANT: Harding, Robert Maxwell  
 ; APPLICANT: Dugdale, Benjamin  
 ; APPLICANT: Beetham, Peter Ronald  
 ; APPLICANT: Hafner, Gregory John  
 ; APPLICANT: Becker, Douglas Kenneth  
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS  
 ; FILE REFERENCE: 09657/002001  
 ; CURRENT APPLICATION NUMBER: US/08/973,068  
 ; CURRENT FILING DATE: 1998-03-12  
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335  
 ; EARLIER FILING DATE: 1996-05-31  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 24  
 ; LENGTH: 547  
 ; TYPE: DNA  
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)  
 US-08-973-068-24

Query Match 58.7%; Score 18.2; DB 3; Length 547;  
 Best Local Similarity 74.2%; Pred. No. 22;  
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagcccccgcgcgc 31  
 DB 366 AGCGCTGGGGTAAATAAGTATAGTCCCGCAGC 336

RESULT 6

```

US-08-973-068-29/c
; Sequence 29, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-29

Query Match          58.7%; Score 18.2; DB 3; Length 622;
Best Local Similarity 74.2%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgct 31
    ||| ||||| ||||| ||||| ||||| |||||
Db 374 AGCAGCGGGGTAATAATAGTCCCCGTGCT 344

RESULT 7
US-08-973-068-27/c
; Sequence 27, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-27

Query Match          58.7%; Score 18.2; DB 3; Length 624;
Best Local Similarity 74.2%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgct 31
    ||| ||||| ||||| ||||| ||||| |||||
Db 375 AGCAGCGGGGTAATAATAGTCCCCGTGCT 345

RESULT 8
US-08-202-186-20/c
; Sequence 20, Application US/08202186
; Patent No. 5756708

```

```

; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1075 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-20

Query Match          58.7%; Score 18.2; DB 1; Length 1075;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgct 31
    ||| ||||| ||||| ||||| ||||| |||||
Db 31 AGCGTGGGGTAATAATAGTCCCCAGCGCT 1

RESULT 9
US-08-973-068-2/c
; Sequence 2, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; CURRENT FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; EARLIER FILING DATE: 1996-05-31
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1075
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: (213)...(737)
US-08-973-068-2

Query Match      58.7%; Score 18.2; DB 3; Length 1075;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtactagcccccgcgct 31
   ||| ||||| ||||| ||| ||||| |||
Db 31 AGCGTGGGGTAATAATAGTCCCGCT 1

RESULT 10
US-08-202-186-23/c
; Sequence 23, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-23

Query Match      58.7%; Score 18.2; DB 1; Length 1089;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtactagcccccgcgct 31
   ||| ||||| ||||| ||| ||||| |||
Db 31 AGCGTGGGGTAATAATAGTCCCGCT 1

RESULT 11
US-08-973-068-8/c
; Sequence 8, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin

```

```

; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; CURRENT FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; EARLIER FILING DATE: 1996-05-31
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 1089
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (281)...(742)
US-08-973-068-8

Query Match      58.7%; Score 18.2; DB 3; Length 1089;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtactagcccccgcgct 31
   ||| ||||| ||||| ||| ||||| |||
Db 31 AGCAGGGGGTAATAATAGTCCCGCT 1

RESULT 12
US-08-202-186-14/c
; Sequence 14, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-14

```

Query Match 58.7%; Score 18.2; DB 1; Length 1110;  
Best Local Similarity 74.2%; Pred. NO. 23;  
Matches 23; Conservative 0; Mismatches 8; Indels 0

**Qy**    1 agccgcggggtaataactagccccgcggct 31  
       |||                      ||| |||  
**Db**    31 AGCGCTGGGGTAATAATAGTCCCCACGCT 1

```

RESULT 13
US-08-208-486-2/c
; Sequence 2, Application US/08208486
; Patent No. 5389531
; GENERAL INFORMATION:
; APPLICANT: Ito, Junetsu
; APPLICANT: Yoo, Seung-ku
; TITLE OF INVENTION: METHODS TO REPLICATE DNA in vitro USING
; TITLE OF INVENTION: PRD1-CATALYZED DNA REPLICATION SYSTEMS
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cahill, Sutton & Thomas
; STREET: 155 Park One, 2141 E. Highland Ave.
; CITY: Phoenix
; STATE: Arizona
; COUNTRY: U.S.A.
; ZIP: 85016

```

```

;
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 1.2 Mb
; COMPUTER: Packard Bell (IBM PC/AT compatible)
; OPERATING SYSTEM: MS-Dos, Version 5.0
; SOFTWARE: WordPerfect Version 5.1
;

```

CURRENT APPLICATION DATA: US/08/208.486  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/869.916  
 FILING DATE: April 14, 1992  
 APPLICATION NUMBER: Japan 240525/91  
 FILING DATE: August 26, 1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Janelle Faunce Raupp  
 REGISTRATION NUMBER: 30,485  
 REFERENCE/DOCKET NUMBER: #3954-A-7  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (602) 956-7000  
 TELEFAX: (602) 495-9475  
 INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2974 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: Double  
 ; TOPOLOGY: Linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; FEATURE: 1-110 E CDS (L-TFR)  
 ; FEATURE: 233-1012 E CDS (PRD1 terminal protein)  
 ; FEATURE: 1016-2677 E CDS (PRD1 DNA polymerase)  
 JS-08-208-486-2

Query Match 57.4%; Score 17.8  
Best Local Similarity 75.9%; Pred. No. 3  
Matches 22; Conservative 0; Mismatch

RESULT 14  
 SS-09-193-792-1/c  
 Sequence 1, Application US/09193792B  
 Patent No. 6180344

```

: GENERAL INFORMATION:
: APPLICANT: Chen, Bin
: TITLE OF INVENTION: 5( Upstream Region Sequences of the MYOD1 Gene
: TITLE OF INVENTION: and Uses Thereof
: FILE REFERENCE: D6015
: CURRENT APPLICATION NUMBER: US/09/193,792B
: CURRENT FILING DATE: 1998-11-17
: PRIOR APPLICATION NUMBER: US 60/065,113
: PRIOR FILING DATE: 1997-11-18
: NUMBER OF SEQ ID NOS: 20
: SEQ ID NO 1
: LENGTH: 2284
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: promoter
: LOCATION: -1537..747
: OTHER INFORMATION: 5( upstream promoter region of the human MYOD1 gene
US-09-193-792-1

```

```
Query Match      55.5%; Score 17.2; DB 4; Length 2284;
Best Local Similarity 73.3%; Pred. No. 64;
Matches 22; Conservative 0; Mismatches 8; Indels 0
```

Qy 1 agcgcgggggtaataactagcccccgcg 30  
||| ||| ||| ||| ||| ||| ||| |||  
Db 2243 AGTCGCCGCTGTAGTCTGCCGCGCGGC 2214

```

RESULT 15
US-08-313-181-1/c
; Sequence 1, Application US/08313181
; Patent No. 5681735
; GENERAL INFORMATION:
; APPLICANT: Emerson, Charles P.
; APPLICANT: Goldhamer, David J.
; TITLE OF INVENTION: Transcription Control Element for
; TITLE OF INVENTION: Increasing Gene Expression in Myoblasts
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street, Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA

```

```

; ZIF: 19103
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patent in Release #1.0, Version #1.30

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,181
FILING DATE: 07-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

```

Search completed: October 24, 2001, 10:00:30  
Job time: 215 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:26 ; Search time 393.52 Seconds  
(without alignments)  
14.360 Million cell updates/sec

Title: US-09-462-955-3

Perfect score: 9

Sequence: 1 taatattac 9

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq:0601:\*

- 1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 9     | 100.0       | 10     | AAF41274 | Yeast NORF gene SA |
| 2          | 9     | 100.0       | 17     | AAV73850 | C. albicans DNA ta |
| 3          | 9     | 100.0       | 18     | AAQ48467 | Nucleotides 970-98 |
| 4          | 9     | 100.0       | 20     | AAQ22027 | C. albicans CS ant |
| 5          | 9     | 100.0       | 20     | AAQ22030 | C. albicans CS ant |
| 6          | 9     | 100.0       | 20     | AAQ22035 | C. albicans CS AUG |
| 7          | 9     | 100.0       | 20     | AAQ41484 | Target region for  |
| 8          | 9     | 100.0       | 20     | AAQ5725  | Human p13 kinase p |
| 9          | 9     | 100.0       | 21     | AAQ22039 | C. albicans CS AUG |
| 10         | 9     | 100.0       | 21     | AAQ06650 | B. anthracis gyrb  |
| 11         | 9     | 100.0       | 21     | AAQ06641 | B. anthracis gyrb  |

|      |   |       |     |    |          |                    |
|------|---|-------|-----|----|----------|--------------------|
| c 12 | 9 | 100.0 | 21  | 20 | AAQ01807 | R. thuringiensis g |
| c 13 | 9 | 100.0 | 24  | 20 | AAQ82184 | Human hyd (h-Hyd)  |
| c 14 | 9 | 100.0 | 24  | 20 | AAQ82184 | Human hyd (h-Hyd)  |
| c 15 | 9 | 100.0 | 25  | 22 | AAQ60809 | S. cerevisiae MET1 |
| c 16 | 9 | 100.0 | 26  | 21 | AAA08601 | PCR primer #4 to a |
| c 17 | 9 | 100.0 | 26  | 21 | AAA08602 | PCR primer #5 to a |
| c 18 | 9 | 100.0 | 26  | 21 | AAZ47072 | Primer SppC to mut |
| c 19 | 9 | 100.0 | 26  | 22 | AAQ31180 | Oligonucleotide Ss |
| c 20 | 9 | 100.0 | 27  | 21 | AAA40331 | Human apolipoprote |
| c 21 | 9 | 100.0 | 34  | 21 | AAZ43142 | Primer sigK-4 to p |
| c 22 | 9 | 100.0 | 35  | 18 | AAQ94383 | Secondary alcohol  |
| c 23 | 9 | 100.0 | 37  | 18 | AAQ95210 | Ligand L13 for per |
| c 24 | 9 | 100.0 | 40  | 17 | AAQ70616 | Sequence of C. tra |
| c 25 | 9 | 100.0 | 42  | 10 | AAQ97079 | Template YC-80 for |
| c 26 | 9 | 100.0 | 49  | 19 | AAV63501 | Staphylococcus aur |
| c 27 | 9 | 100.0 | 50  | 18 | AAV76820 | PCR primer Al-3 us |
| c 28 | 9 | 100.0 | 55  | 20 | AAQ15959 | Oligonucleotide us |
| c 29 | 9 | 100.0 | 57  | 21 | AAQ66232 | peripheral blood m |
| c 30 | 9 | 100.0 | 60  | 20 | AAV83649 | Human secreted pro |
| c 31 | 9 | 100.0 | 72  | 20 | AAQ85096 | Human secreted pro |
| c 32 | 9 | 100.0 | 94  | 21 | AAQ11499 | Human secreted pro |
| c 33 | 9 | 100.0 | 95  | 21 | AAQ28708 | Human secreted pro |
| c 34 | 9 | 100.0 | 97  | 20 | AAQ11476 | Human secreted pro |
| c 35 | 9 | 100.0 | 118 | 21 | AAQ28585 | Sequence 300-7 ide |
| c 36 | 9 | 100.0 | 125 | 21 | AAQ08016 | Human secreted pro |
| c 37 | 9 | 100.0 | 126 | 21 | AAA45603 | Human secreted exp |
| c 38 | 9 | 100.0 | 127 | 21 | AAQ25885 | Human secreted exp |
| c 39 | 9 | 100.0 | 129 | 16 | AAQ21505 | Human secreted pro |
| c 40 | 9 | 100.0 | 129 | 16 | AAQ21505 | Human gene signatu |
| c 41 | 9 | 100.0 | 132 | 21 | AAQ29121 | Human secreted pro |
| c 42 | 9 | 100.0 | 132 | 21 | AAQ43018 | Human secreted exp |
| c 43 | 9 | 100.0 | 134 | 16 | AAQ21752 | Human gene signatu |
| c 44 | 9 | 100.0 | 135 | 18 | AAV77559 | Staphylococcus aur |
| c 45 | 9 | 100.0 | 135 | 21 | AAQ22392 | Human secreted pro |

## ALIGNMENTS

RESULT 1  
AAF41274  
ID AAF41274 standard; DNA; 10 BP.  
XX AAF41274;  
XX  
XX  
DT 23-MAR-2001 (first entry)  
XX  
XX Yeast NORF gene SAGE tag oligonucleotide SEQ ID NO:8013.  
XX  
KW Yeast; Saccharomyces cerevisiae; characterisation; cell cycle; NORF;  
KW nor previously assigned open reading frame; nonannotated ORF; SAGE;  
KW serial analysis of gene expression; antifungal; tag; identification;  
KW linker; PCR primer; ds.  
XX  
XX Saccharomyces cerevisiae.  
XX  
XX WO200077214-A2.  
XX  
PD 21-DEC-2000.  
XX  
XX 14-JUN-2000; 2000WO-US16223.  
XX  
XX 16-JUN-1999; 99US-0335032.  
XX  
XX (UYJO ) UNIV JOHNS HOPKINS.  
XX  
XX Velculescu V, Vogelstein B, Kinzler K;  
XX WPI; 2001-061874/07.  
XX  
XX Yeast gene coding sequences comprising NORF genes with serial analysis  
PT of gene expression (SAGE) tags, useful for studying, monitoring and  
PT affecting phases of the cell cycle.

XX Example; Page 286; 419pp; English.

PS The present invention describes an isolated DNA molecule comprising a

XX coding sequence of a yeast gene selected from a group of 745 NORF (not

CC previously assigned open reading frame; or nonannotated ORF) genes

CC comprising a SAGE (serial analysis of gene expression) tag. Also

CC described are: (1) a method (M1) of using NORF genes to affect the cell

CC cycle comprising administering a NORF gene whose expression varies by at

CC least 10% between any two phases of the cell cycle selected from log

CC phase, S phase and G2/M; (2) a method (M2) for screening candidate

CC antifungal drugs comprising: (a) contacting a test substance with a

CC yeast cell; and (b) monitoring expression of a NORF gene whose

CC expression varies as in M1, where a test substance which modifies the

CC expression of the yeast gene is a candidate antifungal drug; (3) a method

CC (M3) for identifying human genes which are involved in cell cycle

CC progression comprising contacting human DNA with a probe which comprises

CC at least 10 contiguous nucleotides of a NORF gene whose expression varies

CC as in M1; and (4) a method (M4) for identifying a candidate drug as a

CC member of a class of drugs having a characteristic effect on gene

CC expression in a yeast cell comprising contacting a yeast cell with a

CC candidate drug and monitoring expression in the yeast cell of at least 1

CC NORF gene whose expression is affected by the class of drugs. The NORF

CC genes may be used to study, monitor and affect phases of the cell cycle,

CC the differentially expressed genes may be used as markers of phases of

CC the cell cycle. The methods may be used to identify candidate drugs which

CC affect the cell cycle and for identification of antifungal drugs.

CC AAF33268 to AAF44064 represent SAGE tags used in the exemplification of

CC the present invention. AAF33262 to AAF33267 represent linkers and PCR

CC primers used in the SAGE method, in the exemplification of the present

CC invention.

XX Sequence 10 BP; 4 A; 2 C; 0 G; 4 T; 0 other;

SQ

Query Match 100.0%; Score 9; DB 22; Length 10;

Best Local Similarity 100.0%; Pred. No. 9.5e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9

DB 2 taatattac 10

RESULT 2

AAV73850 standard; DNA; 17 BP.

XX AAV73850;

XX 25-FEB-1999 (first entry)

XX C. albicans DNA target region for antisense inhibition.

DE Antisense; inhibition; chiral phosphate linkage; reporter gene; drug;

XX RNase H activity; nuclease resistance; hybridisation; diagnostic;

XX cellular absorption; transport; enzymatic interaction; ss.

XX Candida albicans.

XX US5852188-A.

XX 22-DEC-1998.

XX 19-APR-1996; 96US-0635009.

XX 19-APR-1996; 96US-0635009.

XX 11-JAN-1990; 90US-0463358.

XX 13-AUG-1990; 90US-0566977.

XX 11-JAN-1991; 91WO-US00243.

XX (ISIS-) ISIS PHARM INC.

XX

PI Cook PD;

XX WPI; 1999-080505/07.

XX New oligo-nucleotide(s) for e.g. testing anti-sense activity

PT comprise non-naturally occurring nucleoside unit and chiral

PT phosphate linkages

XX Disclosure; Column 12; 18pp; English.

PS This sequence is used as a target sequence for a novel method to test

XX for antisense activity using an oligonucleotide comprising nucleoside

CC units linked via phosphate linkages in which at least one of the

CC nucleoside units is a non-naturally occurring nucleoside unit and at

CC least two of the nucleoside units are linked via chiral phosphate

CC linkages. The oligonucleotides can be used to test for antisense

CC activity using reporter genes in assays and to test antisense activity

CC against selected cellular target mRNA's in cultured cells. Some of the

CC oligonucleotides are useful for to elicit RNase H activity as a

CC termination event or to increase nuclease resistance. The

CC oligonucleotides are expected to exhibit one or more properties such as

CC hybridisation with target RNA's and DNA's, cellular absorption,

CC transport, or to improve enzymatic interaction without diminishing

CC existing properties giving improved, drugs, diagnostics and research

XX agents.

SQ Sequence 17 BP; 6 A; 3 C; 2 G; 6 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 17;

Best Local Similarity 100.0%; Pred. No. 9e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9

DB 7 taatattac 15

RESULT 3

AAQ48467/C

ID AAQ48467 standard; DNA; 18 BP.

XX AAQ48467;

XX 18-MAR-1994 (first entry)

XX Nucleotides 970-987 of RepA TS mutant gene.

DE Temperature sensitive replication; antibiotic resistance marker gene;

XX site-specific recombination; chromosomal integration; inactivation;

XX heterologous gene expression; thermosensitive plasmid;

XX suicide vector; Lactococcus; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT misc\_difference 3 /\*tag= a

FT /\*note= "wild-type G has been mutated to A

FT at position 972 of ORFA"

FT misc\_difference 8 /\*tag= b

FT /\*note= "wild-type G has been mutated to A

FT at position 977 of ORFA"

FT misc\_difference 11 /\*tag= c

FT /\*note= "wild-type G has been mutated to A

FT at position 980 of ORFA"

FT misc\_difference 18 /\*tag= d

FT /\*note= "wild-type G has been mutated to A

FT at position 987 of ORFA"

XX



PN WO9318164-A.  
 XX 16-SEP-1993.  
 PD 12-MAR-1993; 93WO-FR00248.  
 PF 13-MAR-1992; 92FR-0003034.  
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.  
 XX Gruss A, Maguin E;  
 PI WPI; 1993-303478/38.  
 DR P-PSDB; AAR41755.  
 XX New bacterial plasmid contg. heat sensitive replication system -  
 PT and marker gene, opt..capable of chromosomal integration, used to  
 PT inactivate specific gene or introduce heterologous gene  
 XX Claim 12; Fig 3; 73pp; French.  
 XX Plasmid pWV01 is a cryptic plasmid originally isolated from  
 CC Lactococcus lactis subsp. cremoris; it has a wide host range, being  
 CC able to replicate in both gram-positive and gram-negative bacteria.  
 CC There are 4 open reading frames in pWV01 with ORF-A coding for the  
 CC replication protein RepA which cuts a DNA strand of ORI+. Plasmid  
 CC pGK12 (Appl.Environ.Microbiol., 48; 726 (1984)) was derived from  
 CC pWV01 and contains two antibiotic resistance marker genes.  
 CC Mutagenesis of pGK12 with hydroxylamine resulted in isolation of a  
 CC heat-stable mutant. Sequence analysis localised the mutations to 4  
 CC sites within the RepA coding sequence, i.e. at nucleotides 972,  
 CC 977, 980 and 987 of ORFA (see AAQ48467 for mutant sequence). The  
 CC invention covers plasmids which comprise the temperature sensitive  
 CC ORI; these plasmids can be used to inactivate host genes or to  
 CC introduce heterologous genes into a bacterial host.  
 XX Sequence 18 BP; 9 A; 1 C; 1 G; 7 T; 0 other;  
 SQ

Query Match 100.0%; Score 9; DB 14; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 9e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 taatattac 9  
 Db 14 TAATATTAC 6

RESULT 4  
 AAQ22027  
 ID AAQ22027 standard; DNA; 20 BP.  
 XX AAQ22027;  
 AC AAQ22027;  
 XX 13-JUL-1992 (first entry)  
 DT C. albicans CS antisense oligonucleotide analogue (1287).  
 XX Candida albicans; diagnosis; infection; therapy; CS;  
 KW chitin synthetase; ss.  
 XX Synthetic.  
 OS WO9203455-A.  
 XX WO9203455-A.  
 PN 05-MAR-1992.  
 PD 15-AUG-1991; 91WO-US05802.  
 XX 16-AUG-1990; 90US-0568672.  
 PR (ISIS-) ISIS PHARM INC.  
 PA Hoke GD, Ecker DJ;  
 PI WPI; 1992-096818/12.  
 DR New oligo-nucleotide analogues as anti-sense agents - for the  
 XX treatment and diagnosis of Candida infections  
 PT Example 3; Page 14; 32pp; English.  
 XX A phosphodiester and phosphorothioate analog of this sequence is  
 CC disclosed. The oligonucleotides represented in AAQ22030-41 are tested  
 CC in germ tube assay for inhibition of C. albicans. Oligonucleotide  
 CC 2216, the phosphorothioate oligonucleotide analog of 2214 (AAQ22030),  
 CC hybridisable with mRNA encoding Candida chitin synthetase, showed a  
 CC greater inhibition of Candida germ tube formation relative to the  
 CC other cpds.  
 XX Sequence 20 BP; 6 A; 3 C; 3 G; 8 T; 0 other;  
 SQ

PI Hoke GD, Ecker DJ;  
 XX WPI; 1992-096818/12.  
 DR New oligo-nucleotide analogues as anti-sense agents - for the  
 XX treatment and diagnosis of Candida infections  
 PT Example 1; Page 11; 32pp; English.  
 XX The target RNA for this antisense oligonucleotide is CS  
 CC (initiation of translation). The sequence modulates the  
 CC activity of Candida RNA or DNA and thus can be used for treatment,  
 CC diagnosis and research of Candida infections. Such antisense  
 CC therapies are highly specific and do not cause toxic side effects  
 CC to the host.  
 XX Sequence 20 BP; 6 A; 3 C; 3 G; 8 T; 0 other;  
 SQ

Query Match 100.0%; Score 9; DB 13; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 taatattac 9  
 Db 10 taatattac 18

RESULT 5  
 AAQ22030  
 ID AAQ22030 standard; DNA; 20 BP.  
 XX AAQ22030;  
 AC AAQ22030;  
 XX 13-JUL-1992 (first entry)  
 DT C. albicans CS AUG antisense oligonucleotide analogue (2214).  
 XX Candida albicans; diagnosis; infection; therapy; CS;  
 KW chitin synthetase; phosphodiester; phosphorothioate; ss.  
 XX Synthetic.  
 OS WO9203455-A.  
 XX WO9203455-A.  
 PN 05-MAR-1992.  
 PD 15-AUG-1991; 91WO-US05802.  
 XX 16-AUG-1990; 90US-0568672.  
 PR (ISIS-) ISIS PHARM INC.  
 PA Hoke GD, Ecker DJ;  
 PI WPI; 1992-096818/12.  
 DR New oligo-nucleotide analogues as anti-sense agents - for the  
 XX treatment and diagnosis of Candida infections  
 PT Example 3; Page 14; 32pp; English.  
 XX A phosphodiester and phosphorothioate analog of this sequence is  
 CC disclosed. The oligonucleotides represented in AAQ22030-41 are tested  
 CC in germ tube assay for inhibition of C. albicans. Oligonucleotide  
 CC 2216, the phosphorothioate oligonucleotide analog of 2214 (AAQ22030),  
 CC hybridisable with mRNA encoding Candida chitin synthetase, showed a  
 CC greater inhibition of Candida germ tube formation relative to the  
 CC other cpds.  
 XX Sequence 20 BP; 6 A; 3 C; 3 G; 8 T; 0 other;  
 SQ

```
Query Match      100.0%; Score 9; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.9e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

|    |            |  |                |
|----|------------|--|----------------|
| Qy | 1          | taatattac  | 9              |
|    |            |  |                |
| Db | 10         | taatattac  | 18             |
|    |            |  |                |
|    | RESULT     | 8  |                |
|    | AAC65725/c |  |                |
|    | ID         | AAC65725 standard; DNA; 20 BP.   |                |
| XX | XX         |  |                |
| AC | AAC65725;  |  |                |
| XX |            |  |                |
| XX |            | 16-FEB-2001  | (first entry)  |
| DT |            |  |                |
| XX |            |  |                |
| DE |            | Human p13 kinase p110beta antisense primer                             | SEQ ID NO 36.  |
| XX |            |  |                |
| XX |            | Human; p13 kinase p110 beta; antisense inhibition; primer; cytostatic; |                |
| KW |            | antiinflammatory; antiinfective; ss.                                   |                |
| KW |            |  |                |
| XX |            |  |                |
| XX |            | Homo sapiens.  |                |
| OS |            |  |                |
| XX |            |  |                |
| XX |            | US6133032-A.   |                |
| PN |            |  |                |
| XX |            |  |                |
| XX |            | 17-OCT-2000.   |                |
| PD |            |  |                |
| XX |            |  |                |
| XX |            | 09-SEP-1999;   | 99US-03923350. |
| PF |            |  |                |
| XX |            |  |                |
| XX |            | 09-SEP-1999;   | 99US-03923350. |
| PR |            |  |                |
| XX |            |  |                |
| XX |            | (ISIS-) ISIS PHARM INC.  |                |
| PA |            |  |                |
| PA |            | Monla BP, Cowsert LM;  |                |
| PI |            |  |                |
| XX |            |  |                |
| XX |            |  |                |

Qy 1 taatattac 9  
          |||||  
Db 10 taatattac 18

RESULT 6  
AAQ22035  
ID AAQ22035 standard; DNA; 20 BP.  
XX  
AC AAQ22035;  
XX  
DT 13-JUL-1992 (first entry)  
XX  
DE C. albicans CS AUG antisense oligonucleotide analogue (2938).

XX  
KW Candida albicans; diagnosis; infection; therapy; CS;  
KW chitin synthetase; phosphorothioate; ss.  
KW

|    |   |
|----|---|
| XX | Synthetic.  |
| OS |   |
| XX |   |
| XX | W09203455-A.  |
| PN |   |
| PN |   |
| XX |   |
| XX | 05-MAR-1992.  |
| PD |   |
| XX |   |
| XX | 15-AUG-1991; 91WO-US05802.                                    |
| PF |   |
| PF |   |
| XX |   |
| XX | 16-AUG-1990; 90US-0568672.                                    |
| PR |   |
| XX |   |
| XX | (ISIS-) ISIS PHARM INC..                                      |
| PA |   |
| XX |   |
| XX | Hoke GD, Ecker DJ;  |
| PI |   |
| XX |   |
| XX | WPI; 1992-096818/12.  |
| XX |   |
| XX | New oligo-nucleotide analogues as anti-sense agents - for the |
| PT | treatment and diagnosis of Candida infections                 |
| PT |   |

Example 3; Page 14; 32pp; English.

sequence 20 BP: 6 A; 3 C; 3 G; 8 T; 0 other;

|                       |         |                    |               |            |
|-----------------------|---------|--------------------|---------------|------------|
| Query Match           | 100.0%; | Score 9;           | DB 13;        | Length 20; |
| Best Local Similarity | 100.0%; | Pred. No. 8.9e+03; |               |            |
| Matches               | 9;      | Conservative 0;    | Mismatches 0; | Indels 0;  |
| Gaps                  | 0;      |                    |               |            |

Qy 1 taatattac 9  
| | | | | | | |  
Db 10 taatattac 18

|          |   |
|----------|---|
| RESULT   | 7   |
| AAQ41484 |   |
| ID       | AAQ41484 standard; DNA; 20 BP.              |
| XX       |   |
| XX       |   |
| XX       | AAQ41484;                                   |
| XX       |   |
| DT       | 24-AUG-1993 (first entry)                   |
| XX       |   |
| XX       |   |
| DE       | Target region for <i>Candida albicans</i> . |
| XX       |   |
| XX       |   |

DR WPI: 2000-686014/67.  
XX  
PT Antisense compound 8-30 nucleobases in length targeted to a start codon  
PT of the coding region of human p13 kinase p10beta, useful for  
PT inhibiting the expression of the human polynucleotide -  
XX  
PS Claim 14; Column 61-62; 34pp; English.  
XX  
CC This invention describes a novel antisense compound (I) 8-30 nucleobases  
CC in length targeted to a start codon or nucleobases 4-3174 of the coding  
CC region of human p13 kinase p10beta (II), in which (I) specifically  
CC hybridizes with and inhibits the expression of (II). The products of the  
CC invention have cytostatic, antiinflammatory and antiinfective activity.  
CC (I) is useful for inhibiting the expression of (II) in human cells or  
CC tissues. The antisense compound can be utilized for diagnostics,  
CC therapeutics, prophylaxis and as research reagents and kits. The  
CC antisense compounds may also be useful prophylactically, e.g. to prevent  
CC or delay infection, inflammation or tumor formation. The antisense  
CC compounds are useful for research and diagnostics, because these  
CC compounds hybridize to nucleic acids encoding (II).  
XX  
SQ Sequence 20 BP; 5 A; 3 C; 3 G; 9 T; 0 other;  
  
Query Match 100.0%; Score 9; DB 21; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.9e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 taatattac 9  
Db |||||  
10 TAAATATTAC 2  
  
RESULT 9  
AAQ22039  
ID AAQ22039 standard; DNA; 21 BP.  
AC  
XX  
AC AAQ22039;  
XX  
DT 13-JUL-1992 (first entry)  
XX  
DE C. albicans CS AUG antisense oligonucleotide analogue (3150).  
XX  
KW Candida albicans; diagnosis; infection; therapy; CS;  
KW chitin synthetase; phosphorothioate; ss.  
XX  
OS Synthetic.  
XX  
PN WO9203455-A.  
XX  
PD 05-MAR-1992.  
XX  
PF 15-AUG-1991; 91WO-US05802.  
XX  
PR 16-AUG-1990; 90US-0568672.  
XX  
PA (ISIS-) ISIS PHARM INC.  
XX  
PI Hoke GD, Ecker DJ;  
XX  
WPI: 1992-096818/12.  
XX  
XX New oligo-nucleotide analogues as anti-sense agents - for the  
XX treatment and diagnosis of Candida infections  
XX  
PS Example 3; Page 14; 32pp; English.  
XX  
CC The sequence is a phosphorothioate analog.  
CC The oligonucleotides represented in AAQ22030-41 are tested  
CC in germ tube assay for inhibition of C. albicans. Oligonucleotide  
CC 2216, the phosphorothioate oligonucleotide analog of 2214 (AAQ22030),  
CC hybridisable with mRNA encoding Candida chitin synthetase, showed a  
CC greater inhibition of Candida germ tube formation relative to the

CC other cpds.  
XX  
SQ Sequence 21 BP; 8 A; 4 C; 2 G; 7 T; 0 other;  
  
Query Match 100.0%; Score 9; DB 13; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.8e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 taatattac 9  
Db |||||  
6 taatattac 14  
  
RESULT 10  
AAQ06650/c  
ID AAX06650 standard; DNA; 21 BP.  
XX  
AC AAX06650;  
XX  
DT 08-APR-1999 (first entry)  
XX  
DE B. anthracis gyrB gene specific primer.  
XX  
KW Nucleic acid amplification; detection; Bacillus anthracis; PCR; gyrB;  
KW DNA gyrase subunit B; PCR primer; ss.  
XX  
OS Synthetic.  
OS Bacillus anthracis.  
XX  
PN JP11004693-A.  
XX  
PD 12-JAN-1999.  
XX  
PF 29-SEP-1997; 97JP-0264059.  
XX  
PR 21-APR-1997; 97JP-0102843.  
PR 24-MAR-1997; 97JP-0069900.  
XX  
PA (NIUS ) NIPPON SUISAN KAISHA LTD.  
XX  
WPI: 1999-135786/12.  
XX  
XX New oligonucleotide specific for Bacillus anthracis - amplifies  
XX Bacillus anthracis DNA gyrase subunit B in Bacillus anthracis  
XX  
PS Claim 4; Page 10; 11pp; Japanese.  
XX  
CC The invention relates to an oligonucleotide primer for amplification of  
CC a nucleic-acid target sequence characteristic of Bacillus anthracis. The  
CC oligonucleotide acts as a primer for amplifying the DNA gyrase subunit B  
CC (gyrB) gene for the detection of B. anthracis. A gyrB gene fragment  
CC specific to Bacillus anthracis can be detected using PCR without  
CC extraction of DNA. The method is useful for specific detection of B.  
CC anthracis and distinguishing from B. thuringiensis and B. cereus  
CC strains in a sample. The present sequence represents a primer specific  
CC for the B. anthracis gyrB gene.  
XX  
SQ Sequence 21 BP; 9 A; 3 C; 3 G; 6 T; 0 other;  
  
Query Match 100.0%; Score 9; DB 20; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.8e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 taatattac 9  
Db |||||  
13 TAAATATTAC 5  
  
RESULT 11  
AAX06641/c  
ID AAX06641 standard; DNA; 21 BP.

Thu Oct 25 13:08:45 2001

us-09-462-955-3.rng

```

XX AAX06641;
AC
XX
XX 08-APR-1999 (first entry)
DT
XX
DE B. anthracis gyrB gene specific primer.
XX
XX Nucleic acid amplification; detection; Bacillus cereus; PCR; gyrB;
KW DNA gyrase subunit B; PCR primer; ss.
XX
XX Synthetic.
OS Bacillus anthracis.
XX
XX JP11004691-A.
PN
XX
XX 12-JAN-1999.
PD
XX
XX 29-SEP-1997; 97JP-0264057.
PF
XX
XX 21-APR-1997; 97JP-0102843.
PR
XX
XX 24-MAR-1997; 97JP-0069900.
PA (NIUS ) NIPPON SUISAN KAISHA LTD.
XX
XX WPI; 1999-135785/12.
DR
XX New oligonucleotide specific for Bacillus cereus - amplifies
PT Bacillus cereus DNA gyrase subunit B in Bacillus cereus
XX
XX Examples; Page 10; 11pp; Japanese.
PS
XX The invention relates to an oligonucleotide primer for amplification of
CC a nucleic-acid target sequence characteristic of Bacillus cereus. The
CC oligonucleotide acts as a primer for amplifying the DNA gyrase subunit B
CC (gyrB) gene for the detection of B. cereus. A gyrB gene fragment specific
CC to Bacillus cereus can be detected using PCR without extraction of DNA.
CC The method is useful for specific detection of B. cereus and
CC distinguishing it from B. thuringiensis and B. anthracis strains in
CC a sample. The present sequence represents a primer specific for the
CC B. anthracis gyrB gene.
XX
XX Sequence 21 BP; 9 A; 3 C; 3 G; 6 T; 0 other;
SQ
Query Match 100.0%; Score 9; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 taatattac 9
DB 13 TAATATTAC 5
RESULT 12
AAX01807/c
ID AAX01807 standard; DNA; 21 BP.
XX
XX AAX01807;
XX
XX 09-APR-1999 (first entry)
DT
DE B. thuringiensis gyrB PCR primer #5.
XX
XX gyrB protein; detection; PCR primer; soil; food; ss.
XX
XX Synthetic.
OS Bacillus thuringiensis.
XX
XX JP11004692-A.
PN
XX
XX 12-JAN-1999.
PD
XX
XX 29-SEP-1997; 97JP-0264058.
PF

```

```

XX 21-APR-1997; 97JP-0102843.
PR 24-MAR-1997; 97JP-0069900.
XX
XX (NIUS ) NIPPON SUISAN KAISHA LTD.
XX
XX WPI; 1999-145652/13.
DR
XX Oligonucleotide for Bacillus thuringiensis detection - has nucleic
PT acid sequence obtained from a sequence characteristic for the
PT microorganism
XX
XX Example 2; Page 10; 11pp; Japanese.
PS
XX AAX01803-X01808 are PCR primers used to amplify the Bacillus
CC thuringiensis gyrB genes represented in AAX01800-X01802. These primers
CC can be used in methods to detect the presence of B. thuringiensis in soil
CC or foodstuffs. The primers used in this method can distinguishing
CC between B. thuringiensis, B. cereus and B. anthracis.
XX
XX Sequence 21 BP; 9 A; 3 C; 3 G; 6 T; 0 other;
SQ
Query Match 100.0%; Score 9; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 taatattac 9
DB 13 TAATATTAC 5
RESULT 13
AAX82184
ID AAX82184 standard; DNA; 24 BP.
XX
XX AAX82184;
XX
XX 17-AUG-1999 (first entry)
DT
DE Human hyd (h-Hyd) protein DNA amplifying primer.
XX
XX Human; hyd protein; h-Hyd; hect domain; E6AP; cell growth; heart;
KW intracellular protein decomposition; sperm cell; PCR primer; ss.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX JP11146786-A.
PN
XX
XX 02-JUN-1999.
PD
XX
XX 17-NOV-1997; 97JP-0314935.
PF
XX
XX 17-NOV-1997; 97JP-0314935.
PR
XX (SUME ) SUMITOMO ELECTRIC IND CO.
XX
XX WPI; 1999-378996/32.
XX
XX New h-Hyd protein - and DNA encoding it
PT
XX Examples; Page 5; 20pp; Japanese.
PS
XX The invention relates to a human-hyd (h-Hyd) protein having hect domains
CC and homology to the C-terminal domain of E6AP. h-Hyd is presumed to
CC participate in intracellular protein decomposition and cell growth
CC especially in heart and sperm cells. Sequences AAX82184-187 represent
CC PCR primers for amplifying the h-Hyd protein DNA.
XX
XX Sequence 24 BP; 11 A; 5 C; 2 G; 6 T; 0 other;
SQ

```

```

Query Match      100.0%; Score 9; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 4 taatattac 12

RESULT 14
ID AAX82184/c
XX AAX82184 standard; DNA; 24 BP.
AC AAX82184;
XX 17-AUG-1999 (first entry)
XX Human hyd (h-Hyd) protein DNA amplifying primer.
XX Human: hyd protein; h-Hyd; hect domain; E6AP; cell growth; heart;
KW intracellular protein decomposition; sperm cell; PCR primer; ss.
XX Synthetic.
OS Homo sapiens.
XX JP11146786-A.
XX 02-JUN-1999.
XX 17-NOV-1997; 97JP-0314935.
XX 17-NOV-1997; 97JP-0314935.
XX (SUME ) SUMITOMO ELECTRIC IND CO.
XX WPI; 1999-378996/32.
XX New h-Hyd protein - and DNA encoding it
XX Examples; Page 5; 20pp; Japanese.
XX The invention relates to a human-hyd (h-Hyd) protein having hect domains
CC and homology to the C-terminal domain of E6AP. h-Hyd is presumed to
CC participate in intracellular protein decomposition and cell growth
CC especially in heart and sperm cells. Sequences AAX82184-187 represent
CC PCR primers for amplifying the h-Hyd protein DNA.
XX Sequence 24 BP; 11 A; 5 C; 2 G; 6 T; 0 other;

Query Match      100.0%; Score 9; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 11 TAATATTAC 3

RESULT 15
ID AAF60809
XX AAF60809 standard; DNA; 25 BP.
XX AAF60809;
XX 04-MAY-2001 (first entry)
XX S. cerevisiae MET14 PCR primer SEQ ID 11.
XX Microorganism; sulfite production; sulfite cycle; food production; wine;
KW beer; taste stabilizer; non-volatile complex; carbonyl compound; yeast;
KW oxidation; fermentation; PCR primer; ss.
XX

```

```

OS Saccharomyces cerevisiae.
XX DE19923950-A1.
PN 25-JAN-2001.
XX 25-MAY-1999; 99DE-1023950.
PF 25-MAY-1999; 99DE-1023950.
PR (STAHL/) STAHL U.
XX Donalis U, Stahl U;
PI WPI; 2001-148153/16.
XX New microorganisms that produce high sulfite levels at a late stage in
PT their growth, useful for producing beer, prevent development of
PT off-flavors by oxidation
XX Example 4; Page 8; 14pp; German.
XX This invention describes novel microorganisms (A) able to produce delayed
CC and large amounts of sulfite. The microorganisms comprise a DNA construct
CC (I) containing one or more genes (II) involved in the sulfite cycle under
CC the control of a promoter. The high sulfite concentration appears at a
CC late stage of substrate utilization, in the stationary growth phase, in
CC 60-90% of that achieved in the growth phase. (A), particularly bacteria
CC and/or yeast, are used for the production of foods, wine, beer or desired
CC metabolic end products. Sulfite stabilizes the taste of beer by forming
CC non-volatile complexes with carbonyl compounds (formed by oxidation and
CC responsible for off-flavors) and by preventing oxidation (reducing
CC agent). (A) produce significant amounts of sulfite only at a late stage
CC in its growth, after the fermentation product has formed, avoiding
CC premature formation of complexes and eliminating the need to add sulfite
CC to the finished product.
XX Sequence 25 BP; 7 A; 7 C; 3 G; 8 T; 0 other;

Query Match      100.0%; Score 9; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 14 taatattac 22

Search completed: October 24, 2001, 10:07:27
Job time: 632 sec

```

us-09-462-955-3.rng

---

Thu Oct 25 13:08:45 2001

•  
•  
•  
•  
•

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:52 ; Search time 5479.82 Seconds  
(without alignments)  
15.525 Million cell updates/sec

Title: US-09-462-955-3  
Perfect score: 9  
Sequence: 1 taatattac 9

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estomi:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estrol:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*





SUMMARIES

| Result No. | Score | Query Match | Length | DB  | ID       | Description        |
|------------|-------|-------------|--------|-----|----------|--------------------|
| 1          | 9     | 100.0       | 24     | 246 | AZ595661 | AZ595661 1M0408K24 |
| 2          | 9     | 100.0       | 25     | 249 | AZ803817 | AZ803817 2M0064H22 |
| 3          | 9     | 100.0       | 25     | 246 | AZ595666 | AZ595666 1M0408L23 |
| 4          | 9     | 100.0       | 31     | 244 | AZ485621 | AZ485621 1M0313H08 |
| 5          | 9     | 100.0       | 34     | 24  | AI784676 | AI784676 ta03f03.x |
| 6          | 9     | 100.0       | 34     | 246 | AZ589161 | AZ589161 1M0397O22 |
| 7          | 9     | 100.0       | 36     | 244 | AZ474278 | AZ474278 1M0290F15 |
| 8          | 9     | 100.0       | 51     | 106 | AU007795 | AU007795 AU007795  |
| 9          | 9     | 100.0       | 58     | 4   | AA284056 | AA284056 zs60d05.s |
| 10         | 9     | 100.0       | 58     | 121 | AW827116 | AW827116 xn13e07.y |
| 11         | 9     | 100.0       | 64     | 244 | AZ471237 | AZ471237 1M0285C13 |
| 12         | 9     | 100.0       | 66     | 246 | AZ587595 | AZ587595 1M0395E24 |
| 13         | 9     | 100.0       | 68     | 164 | BE191328 | BE191328 sn71b06.y |
| 14         | 9     | 100.0       | 69     | 249 | AZ755547 | AZ755547 ev01c06.x |
| 15         | 9     | 100.0       | 71     | 174 | BG226871 | BG226871 kp99a04.y |
| 16         | 9     | 100.0       | 72     | 256 | AZ292450 | AZ292450 479.dif26 |
| 17         | 9     | 100.0       | 73     | 102 | AI833171 | AI833171 at65e03.x |
| 18         | 9     | 100.0       | 73     | 102 | AI833171 | AI833171 at65e03.x |
| 19         | 9     | 100.0       | 73     | 118 | AW600128 | AW600128 SWL4CAK10 |
| 20         | 9     | 100.0       | 73     | 118 | AW600148 | AW600148 SWL4CAK10 |
| 21         | 9     | 100.0       | 73     | 145 | BF145150 | BF145150 SWOV13CAN |
| 22         | 9     | 100.0       | 81     | 156 | C11376   | C11376 C11376.Yu1  |
| 23         | 9     | 100.0       | 81     | 243 | AZ427672 | AZ427672 1M0209G15 |
| 24         | 9     | 100.0       | 82     | 244 | AZ489440 | AZ489440 1M0321L10 |
| 25         | 9     | 100.0       | 84     | 162 | BE022876 | BE022876 sn77f04.y |
| 26         | 9     | 100.0       | 84     | 249 | AZ783162 | AZ783162 2M0024D08 |
| 27         | 9     | 100.0       | 85     | 4   | AA273976 | AA273976 vb94a07.r |
| 28         | 9     | 100.0       | 85     | 7   | AA445562 | AA445562 vf61h07.r |
| 29         | 9     | 100.0       | 86     | 120 | AW765103 | AW765103 ga88b07.y |
| 30         | 9     | 100.0       | 87     | 13  | AA865195 | AA865195 oh28c05.s |
| 31         | 9     | 100.0       | 87     | 241 | AZ318000 | AZ318000 1M0036K24 |
| 32         | 9     | 100.0       | 87     | 4   | AA276053 | AA276053 vc36h08.r |
| 33         | 9     | 100.0       | 88     | 162 | BE057653 | BE057653 sn05c09.y |
| 34         | 9     | 100.0       | 88     | 244 | AZ494555 | AZ494555 1M0329H20 |
| 35         | 9     | 100.0       | 88     | 246 | AZ602035 | AZ602035 1M0420A07 |
| 36         | 9     | 100.0       | 89     | 13  | AA869288 | AA869288 vq50all.r |
| 37         | 9     | 100.0       | 91     | 13  | AA910700 | AA910700 ok89e10.s |
| 38         | 9     | 100.0       | 94     | 11  | AA708850 | AA708850 z158b03.s |
| 39         | 9     | 100.0       | 94     | 12  | AA841919 | AA841919 MB3D6V8H1 |
| 40         | 9     | 100.0       | 94     | 18  | AI321285 | AI321285 d6e03nm.r |
| 41         | 9     | 100.0       | 94     | 241 | AZ322497 | AZ322497 1M0043008 |
| 42         | 9     | 100.0       | 97     | 8   | AA498371 | AA498371 vh43c01.r |
| 43         | 9     | 100.0       | 97     | 8   | AA498371 | AA498371 vh43c01.r |
| 44         | 9     | 100.0       | 97     | 111 | AW099103 | AW099103 sd34e03.y |
| 45         | 9     | 100.0       | 97     | 122 | AW909872 | AW909872 ur87d02.x |

ALIGNMENTS

|            |  |                    |                |                      |
|------------|--|--------------------|----------------|----------------------|
| RESULT 1   | AZ595661   | 24 bp              | DNA            | 13-DEC-2000          |
| LOCUS      | 1M0408K24F   | Mouse 10kb plasmid | UUGCLM library | Mus musculus genomic |
| DEFINITION | clone UUGCLM0408K24 F, DNA sequence.   |                    |                |                      |
| ACCESSION  | AZ595661.1   | GI:11717851        |                |                      |
| VERSION    | GSS.   |                    |                |                      |
| KEYWORDS   | house mouse.   |                    |                |                      |
| SOURCE     | Mus musculus   |                    |                |                      |
| ORGANISM   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.   |                    |                |                      |
| REFERENCE  | 1 (bases 1 to 24)  |                    |                |                      |
| AUTHORS    | Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R. |                    |                |                      |

TITLE  
JOURNAL  
COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0408 row: K column: 24  
Seq primer: CGTGTAAACAGCGGCAGT  
Class: plasmid ends  
High quality sequence stop: 24.  
Location/Qualifiers  
1. .24

FEATURES  
source

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCLM0408K24"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil47321141gblAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

6 a 3 c 0 g 15 t  
Query Match 100.0%; Score 9; DB 246; Length 24;  
Best Local Similarity 100.0%; Pred. No. 7.3e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
Db 3 TAATATTAC 11

RESULT 2  
AZ803817/c

LOCUS  
DEFINITION  
2M0064H22F Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
clone UUGCLM0064H22 F, DNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AZ803817 25 bp DNA  
GSS  
16-FEB-2001  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 25)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0408 row: L column: 23  
 Seq primer: CGTTGTAACGACGCGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 28.  
 Location/Qualifiers

1. .28  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0408L23"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 12 a 3 c 2 g 11 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 246; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 7.3e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9  
 | | | | | | | | | |  
 Db 28 TAATATTAC 20

RESULT 4  
 AZ485621 31 bp DNA GSS 05-OCT-2000  
 LOCUS  
 DEFINITION  
 1M0313H08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0313H08 F, DNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 31)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0064 row: H column: 22  
 Seq primer: CGTTGTAACGACGCGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 25.  
 Location/Qualifiers

1. .25  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0064H22"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 9 a 6 c 3 g 7 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 249; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 7.3e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9  
 | | | | | | | | | |  
 Db 9 TAATATTAC 1

RESULT 3  
 AZ595666/c 28 bp DNA GSS 13-DEC-2000  
 LOCUS  
 DEFINITION  
 1M0408L23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0408L23 F, DNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 28)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly



Thu Oct 25 13:08:48 2001

Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 36.  
 Location/Qualifiers

FEATURES  
 source

1. 36  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0290F15"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 13 a 8 c 4 g 9 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 246; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 7.2e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taattattac 9  
 |||||

Db 6 TAATATTAC 14

RESULT 7  
 AZ474278  
 LOCUS  
 DEFINITION  
 1M0290F15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0290F15 R, DNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 36)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00

BASE COUNT 14 a 7 c 6 g 9 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 244; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 7.2e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taattattac 9  
 |||||

Db 11 TAATATTAC 19

RESULT 8  
 AU007795/c  
 LOCUS  
 DEFINITION  
 AU007795 Schizosaccharomyces pombe late log phase cDNA  
 Schizosaccharomyces pombe cDNA clone spc02513, mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Schizosaccharomyces pombe  
 Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 Schizosaccharomycetales; Schizosaccharomycetaceae;  
 Schizosaccharomycetes.  
 1 (bases 1 to 51)  
 Morimyo,M. and Mita,K.  
 Identification of expressed sequence tags of Schizosaccharomycetes  
 pombe  
 Unpublished (1998)  
 Contact: Mitsuoki Morimyo  
 Genome Research Group  
 National Institute of Radiological Sciences  
 9-1, Anagawa-4-chome, Inage-ku, Chiba 263-8555, Japan  
 Email: morimyo@nirs.go.jp.  
 Location/Qualifiers

FEATURES  
 source

1. 51  
 /organism="Schizosaccharomyces pombe"  
 /strain="972"  
 /db\_xref="taxon:4896"

```

/clone="spc02513"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mp19 DNA and the direction of DNA
sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT      22 a      3 c      7 g      19 t
ORIGIN

```

```

Query Match      100.0%; Score 9; DB 106; Length 51;
Best Local Similarity 100.0%; Pred. No. 7.1e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 17 TAATATTAC 9

```

```

RESULT 9
AA284056      58 bp      mRNA      EST      15-DEC-1997
LOCUS      ZS60d05.s1 Stratagene schizo brain s11 Homo sapiens cDNA clone
DEFINITION      IMAGE:701865 3', mRNA sequence.
ACCESSION      AA284056
VERSION      AA284056.1 GI:1928337
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 58)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,
T., Waterston, R., and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -41ml3 fwd. ET from Amersham.

```

```

FEATURES
source
1..58
/organism="Homo sapiens"
/db_xref="GDB:5742536"
/db_xref="taxon:9606"
/clone="IMAGE:701865"
/clone_lib="Stratagene schizo brain s11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site1: EcoRI; Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric
Diseases (Unpublished) Stanley Neurovirology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT      22 a      5 c      16 g      15 t

```

## ORIGIN

```

Query Match      100.0%; Score 9; DB 4; Length 58;
Best Local Similarity 100.0%; Pred. No. 7.1e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 8 TAATATTAC 16

```

```

RESULT 10
AW827116      58 bp      mRNA      EST      18-MAY-2000
LOCUS      xn13e07.y1 NCI_CGAP_Li5 Homo sapiens cDNA clone IMAGE:2693604 5',
DEFINITION      mRNA sequence.
ACCESSION      AW827116
VERSION      AW827116.1 GI:7920970
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 58)
NCI-CGAP http://www.ncbi.nlm.nih.gov/hcicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Putative full length read
The vector to vector length is 59
Seq primer: -40RP from Gibco.

```

## FEATURES source

```

1..58
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2693604"
/clone_lib="NCI_CGAP_Li5"
/tissue_type="hepatic adenoma"
/lab_host="DH10B"
/note="Organ: liver; Vector: pCMV-SPORT4; Site1: Salt;
Site2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 0.8 kb."
BASE COUNT      24 a      10 c      7 g      17 t
ORIGIN

```

```

Query Match      100.0%; Score 9; DB 121; Length 58;
Best Local Similarity 100.0%; Pred. No. 7.1e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 8 TAATATTAC 16

```

```

RESULT 11
AAZ471237/c      64 bp      DNA      GSS      04-OCT-2000
LOCUS      1M0285C13R Mouse 10kb plasmid U06C1M library Mus musculus genomic
DEFINITION      clone U06C1M0285C13 R, DNA sequence.
ACCESSION      AZ471237
VERSION      AZ471237.1 GI:10629362

```

Thu Oct 25 13:08:48 2001

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .64
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0285C13"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g1147321141gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT
17 a 1 c 19 g 27 t
ORIGIN
Query Match 100.0%; Score 9; DB 244; Length 64;
Best Local Similarity 100.0%; Pred. No. 7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taattattac 9
Db 50 TAATATTAC 42
RESULT 12
LOCUS
DEFINITION
ACCESSION
A2587595
1M0395E24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0395E24 F, DNA sequence.
A2587595

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .66
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0395E24"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g1147321141gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT
34 a 11 c 3 g 18 t
ORIGIN
Query Match 100.0%; Score 9; DB 246; Length 66;
Best Local Similarity 100.0%; Pred. No. 7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taattattac 9
Db 40 TAATATTAC 48
RESULT 13
LOCUS
DEFINITION
ACCESSION
A2587595
1M0395E24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0395E24 F, DNA sequence.
A2587595

```

```

ACCESSION BE191328
VERSION BE191328.1 GI:8670221
KEYWORDS EST
SOURCE Glycine max
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 68)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
, Y., Person, B., Swaller, F., Gibbons, M., Pape, D., Harvey, N., Schurk
, R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
, R., Waterston, R. and Willson, R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com.
FEATURES
source
1..68
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl038-12"
/clone_lib="Gm-cl038"
/tissue_type="Senescing leaves, greenhouse grown"
/dev_stage="mature plants"
/lab_host="DH10B"
/notes="Vector: pT73Pac (pT73, Pharmacia); Site_1: EcoRI;
Site_2: HindIII; This cDNA library was constructed from
mRNA isolated from the senescing leaf tissue of mature
greenhouse grown plants of the cultivar Williams 82.
Complementary DNA was synthesized from mRNA using a 3'
anchored poly(dT) primer. EcoRI adapters were ligated to
the blunt ended cDNA fragments followed by digestion with
EcoRI and HindIII. The cDNA fragments were directionally
cloned into the EcoRI-HindIII restriction site of the
pT73-Pac vector. The ligated cDNA fragments were
transformed into DH10B host cells (Gibco BRL). This
library was constructed by Dr. Randy Shoemaker and Dr.
John Erpelting."
BASE COUNT 15 a 12 c 10 g 31 t
ORIGIN
Query Match 100.0%; Score 9; DB 164; Length 68;
Best Local Similarity 100.0%; Pred. No. 7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 taattattac 9
Db 44 TAATATTAC 36
RESULT 14
A2755547/c 69 bp DNA GSS 01-MAR-2001
LOCUS ev01c06.x1 PAX3 CASTING Library 'ev' Homo sapiens genomic clone
DEFINITION ev01c06 random, DNA sequence.
ACCESSION A2755547
VERSION A2755547
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 69)
Barber, T.D., Barber, M.C., Tomescu, O., Barr, F., Ruben, S. and
Friedman, R.B.
TITLE Cyclic amplification and selection of target genes regulated by
Pax3 and PAX3/FKHR in embryogenesis and alveolar rhabdomyosarcoma
JOURNAL Unpublished (2000)
COMMENT Contact: Friedman TB
Laboratory of Molecular Genetics
National Institute on Deafness and Other Communication Disorders,
National Institutes of Health
5 Research Court, Room 2A-15, Rockville, MD 20850, USA
Tel: 301 402 7580
Fax: 301 496 7882
Email: friedman@nidcd.nih.gov
Plate: 01 row: c column: 06
Seq primer: -21M13 forward primer (ABI)
Class: random plasmid subclone.
FEATURES
source
1..69
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ev01c06"
/clone_lib="PAX3 CASTING Library 'ev'"
/sex="Male"
/lab_host="DH10B"
/notes="Vector: pGEM-T Easy; Human genomic DNA was
partially digested with Sau3AI, ligated to ds linkers,
and enriched for binding to human PAX3dQ+ protein using a
Whole Genome PCR-based strategy. DNA fragments containing
putative PAX3dQ+ binding sites were amplified by PCR and
cloned into pGEM-T Easy (Promega). The ligation products
were transformed into DH10B electrocompetent cells (Life
Technologies)."
BASE COUNT 18 a 13 c 11 g 27 t
ORIGIN
Query Match 100.0%; Score 9; DB 249; Length 69;
Best Local Similarity 100.0%; Pred. No. 7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 taattattac 9
Db 48 TAATATTAC 40
RESULT 15
BG226871 71 bp mRNA EST 05-APR-2001
LOCUS kp99a04.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5', mRNA
DEFINITION sequence.
ACCESSION BG226871
VERSION BG226871.1 GI:12714426
KEYWORDS EST.
SOURCE Strongyloides stercoralis.
ORGANISM Strongyloides stercoralis
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Strongyloidea.
REFERENCE 1 (bases 1 to 71)
AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wyllie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
Gibbons, M., Rutter, E., Bennett, J., Franklin, C., Tsagaris, R.,
Runko, L., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
Willson, R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999

```

Thu Oct 25 13:08:48 2001

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
The library was constructed by Dr. Thomas Nutman and colleagues of  
NIAID, NIH (tnutman@nih.gov). DNA sequencing by: Washington  
University Genome Sequencing Center St. Louis.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 62.

# FEATURES

source

1. .71  
/organism="Strongyloides stercoralis"  
/strain="Rhabditiform larvae obtained from gerbils"  
/db\_xref="taxon:6248"  
/clone\_lib="MBN95TW-SSR"  
/lab\_host="XL-1 Blue MRF" (Stratagene)  
/note="vector: Lambda Uni-ZAP XR (Stratagene); Site\_1:  
EcoRI; Site\_2: XhoI; mRNA was purified from 2 x 10E3  
rhabditiform larvae which had been isolated from gerbils  
experimentally infected with larvae originally isolated  
from experimentally infected dogs. cDNA was constructed  
and using adaptors, was cloned unidirectionally into the  
vector from the EcoRI site to the XhoI site. The library  
has an unamplified titer of 1 x 10E3 pfu/ml and an  
amplified, undiluted titer of 9 x 10E11 pfu/ml. The  
average insert size of the unamplified library is 675 bp  
(range, 100-1700)."

BASE COUNT 35 a 9 c 6 g 21 t  
ORIGIN

Query Match 100.0%; Score 9; DB 174; Length 71;  
Best Local Similarity 100.0%; Pred. No. 7e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 taatattac 9  
| | | | | | | |  
Db 62 TAATATTAC 70

Search completed: October 24, 2001, 13:13:56  
Job time: 11821 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:30 ; Search time 180.6 Seconds  
(without alignments)  
9,434 Million cell updates/sec

Title: US-09-462-955-3

Perfect score: 9

Sequence: 1 taatattac 9

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A-COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B-COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS-COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 9     | 100.0       | 17     | 1  | US-08-058-023-3   |
| 2          | 9     | 100.0       | 20     | 1  | US-08-297-703-3   |
| 3          | 9     | 100.0       | 20     | 1  | US-08-229-145-1   |
| 4          | 9     | 100.0       | 20     | 1  | US-08-229-145-6   |
| 5          | 9     | 100.0       | 20     | 3  | US-09-392-350-36  |
| 6          | 9     | 100.0       | 20     | 5  | PCT-US91-05802-1  |
| 7          | 9     | 100.0       | 20     | 5  | PCT-US91-05802-6  |
| 8          | 9     | 100.0       | 21     | 1  | US-08-229-145-10  |
| 9          | 9     | 100.0       | 21     | 3  | US-09-046-578-11  |
| 10         | 9     | 100.0       | 21     | 5  | PCT-US91-05802-10 |
| 11         | 9     | 100.0       | 34     | 4  | US-08-793-701-45  |
| 12         | 9     | 100.0       | 35     | 3  | US-09-051-914-10  |
| 13         | 9     | 100.0       | 37     | 2  | US-08-805-191-7   |
| 14         | 9     | 100.0       | 72     | 1  | US-08-433-126A-39 |
| 15         | 9     | 100.0       | 72     | 1  | US-08-433-124A-39 |
| 16         | 9     | 100.0       | 72     | 3  | US-08-976-413A-39 |
| 17         | 9     | 100.0       | 72     | 5  | PCT-US96-06059-39 |
| 18         | 9     | 100.0       | 102    | 4  | US-09-276-531-6   |
| 19         | 9     | 100.0       | 166    | 4  | US-08-905-223-107 |
| 20         | 9     | 100.0       | 230    | 2  | US-08-687-080-106 |
| 21         | 9     | 100.0       | 253    | 6  | 5242821-10        |
| 22         | 9     | 100.0       | 266    | 2  | US-08-353-476-60  |
| 23         | 9     | 100.0       | 311    | 4  | US-08-916-576B-40 |
| 24         | 9     | 100.0       | 311    | 4  | US-08-916-576B-40 |
| 25         | 9     | 100.0       | 351    | 1  | US-08-281-916-7   |
| 26         | 9     | 100.0       | 351    | 1  | US-08-281-916-8   |
| 27         | 9     | 100.0       | 351    | 2  | US-08-460-725-9   |

|    |   |       |     |   |                   |                   |
|----|---|-------|-----|---|-------------------|-------------------|
| 28 | 9 | 100.0 | 351 | 2 | US-08-460-725-10  | Sequence 10, Appl |
| 29 | 9 | 100.0 | 358 | 1 | US-08-686-878A-3  | Sequence 3, Appl  |
| 30 | 9 | 100.0 | 399 | 1 | US-07-885-970A-13 | Sequence 13, Appl |
| 31 | 9 | 100.0 | 399 | 1 | US-08-298-587A-13 | Sequence 13, Appl |
| 32 | 9 | 100.0 | 399 | 1 | US-08-530-797-12  | Sequence 12, Appl |
| 33 | 9 | 100.0 | 399 | 1 | US-08-298-829-13  | Sequence 12, Appl |
| 34 | 9 | 100.0 | 399 | 2 | US-08-787-335-12  | Sequence 12, Appl |
| 35 | 9 | 100.0 | 401 | 4 | US-09-276-531-59  | Sequence 59, Appl |
| 36 | 9 | 100.0 | 405 | 1 | US-08-307-499-45  | Sequence 45, Appl |
| 37 | 9 | 100.0 | 405 | 4 | US-09-299-268-45  | Sequence 45, Appl |
| 38 | 9 | 100.0 | 485 | 3 | US-09-284-782-17  | Sequence 17, Appl |
| 39 | 9 | 100.0 | 495 | 3 | US-08-691-563C-59 | Sequence 59, Appl |
| 40 | 9 | 100.0 | 513 | 1 | US-08-650-528-4   | Sequence 4, Appl  |
| 41 | 9 | 100.0 | 513 | 3 | US-09-060-584-4   | Sequence 4, Appl  |
| 42 | 9 | 100.0 | 513 | 3 | US-09-413-140A-4  | Sequence 4, Appl  |
| 43 | 9 | 100.0 | 516 | 1 | US-08-280-320-1   | Sequence 1, Appl  |
| 44 | 9 | 100.0 | 551 | 4 | US-08-998-416-86  | Sequence 86, Appl |
| 45 | 9 | 100.0 | 551 | 4 | US-08-998-416-86  | Sequence 86, Appl |

## ALIGNMENTS

RESULT 1  
US-08-058-023-3  
; Sequence 3, Application US/08058023  
; Patent No. 5521302  
; GENERAL INFORMATION:  
; APPLICANT: Cook, Phillip D.  
; TITLE OF INVENTION: OLIGONUCLEOTIDES HAVING CHIRAL  
; TITLE OF INVENTION: PHOSPHORUS LINKAGES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz  
; ADDRESSEE: and No. 5521302ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.  
; ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch disk, 720 Kb  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/058,023  
FILING DATE: 05-MAY-1993  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Joseph Luccl  
REGISTRATION NUMBER: 33,307  
REFERENCE/DOCKET NUMBER: ISIS-1053  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-058-023-3

Query Match 100.0%; Score 9; DB 1; Length 17;  
Best Local Similarity 100.0%; Pred. No. 2.5e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 taatattac 9

Thu Oct 25 13:08:47 2001

us-09-462-955-3.rni

```

Db      7 TAATATTAC 15

RESULT 2
US-08-297-703-3
; Sequence 3, Application US/08297703
; Patent No. 5506212
; GENERAL INFORMATION:
; APPLICANT: Hake,, Glenn
; TITLE OF INVENTION: Stereoisomerically Pure
; TITLE OF INVENTION: Phosphorothioate
; TITLE OF INVENTION: Oligonucleotides
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John W. Caldwell
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297.703
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/777,007
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Caldwell, John W.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: ISIS-0015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-368-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-297-703-3

Query Match      100.0%; Score 9; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 taatattac 9
Db      10 TAATATTAC 18

RESULT 3
US-08-229-145-1
; Sequence 1, Application US/08229145
; Patent No. 5691461
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn D.
; TITLE OF INVENTION: INHIBITION OF CANDIDA USING
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/229,145
; FILING DATE: Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: ISIS-1421
; TELECOMMUNICATION INFORMATION:
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/229,145
; FILING DATE: Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: ISIS-1421
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 20 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cdna  
 HYPOTHETICAL: NO  
 ANTI-SENSE: YES  
 US-08-229-145-6

Query Match 100.0%; Score 9; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
 |||||  
 Db 10 TAATATTAC 18

## RESULT 5

US-09-392-350-36/c  
 Sequence 36, Application US/09392350  
 Patent No. 6133032  
 GENERAL INFORMATION:  
 APPLICANT: Brett P. Monia  
 APPLICANT: Lex M. Cowsett  
 TITLE OF INVENTION: ANTISENSE MODULATION OF P13 KINASE P110 BETA EXPRESSION  
 FILE REFERENCE: RFS-0075  
 CURRENT APPLICATION NUMBER: US/09/392,350  
 CURRENT FILING DATE: 1999-09-08  
 NUMBER OF SEQ ID NOS: 47  
 SEQ ID NO 36  
 LENGTH: 20  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Antisense Oligonucleotide  
 US-09-392-350-36

Query Match 100.0%; Score 9; DB 3; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
 |||||  
 Db 10 TAATATTAC 2

## RESULT 6

PCT-US91-05802-1  
 Sequence 1, Application PC/TUS9105802  
 GENERAL INFORMATION:  
 APPLICANT: Hoke, Glenn C.  
 APPLICANT: Ecker, David J.  
 TITLE OF INVENTION: Inhibition of Candida  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz  
 ADDRESS: Norris  
 STREET: One Liberty Place-- 46th floor  
 CITY: Philadelphia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb  
 MEDIUM TYPE: STORAGE  
 COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WORDPERFECT 5.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US91/05802  
 FILING DATE: 19910815  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Licata, Jane M.  
 REGISTRATION NUMBER: 32,257  
 REFERENCE/DOCKET NUMBER: ISIS-0432  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 20 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Other nucleic acid  
 HYPOTHETICAL: NO  
 ANTI-SENSE: YES  
 PCT-US91-05802-1

Query Match 100.0%; Score 9; DB 5; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9  
 |||||  
 Db 10 TAATATTAC 18

## RESULT 7

PCT-US91-05802-6  
 Sequence 6, Application PC/TUS9105802  
 GENERAL INFORMATION:  
 APPLICANT: Hoke, Glenn C.  
 APPLICANT: Ecker, David J.  
 TITLE OF INVENTION: Inhibition of Candida  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz  
 ADDRESS: Norris  
 STREET: One Liberty Place-- 46th floor  
 CITY: Philadelphia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb  
 MEDIUM TYPE: STORAGE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WORDPERFECT 5.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US91/05802  
 FILING DATE: 19910815  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Licata, Jane M.  
 REGISTRATION NUMBER: 32,257  
 REFERENCE/DOCKET NUMBER: ISIS-0432  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 20 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear

```
; MOLECULE TYPE: Other nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
PCT-US91-05802-6

Query Match      100.0%; Score 9; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
   |||||||
DB 10 TAATATTAC 18

RESULT 8
US-08-229-145-10
; Sequence 10, Application US/08229145
; Patent No. 5691461
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn D.
; APPLICANT: Ecker, David J.
; TITLE OF INVENTION: INHIBITION OF CANDIDA USING OLIGONUCLEOTIDES
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 5691461ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/229,145
; FILING DATE: Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: ISIS-1421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-229-145-10

Query Match      100.0%; Score 9; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
   |||||||
DB 6 TAATATTAC 14

RESULT 9
US-09-046-578-11/c
; Sequence 11, Application US/09046578
```

```
; Patent No. 6087104
; GENERAL INFORMATION:
; APPLICANT: Yamada, Shoichi
; APPLICANT: Venkateswaran, Kasthuri
; APPLICANT: Ohashi, Ei-ji
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR DETECTION OF
; TITLE OF INVENTION: BACILLUS CEREUS GROUP BACTERIA HARMFUL TO MAMMALS, AND THE
; TITLE OF INVENTION: METHOD FOR DETECTION WITH THE OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARMSTRONG, WESTERMAN, HATTORI, McLELAND &
; ADDRESSEE: NAUGHTON
; STREET: 1725 K Street, NW Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,578
; FILING DATE: 24-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mclelland, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 980313
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-659-2930
; TELEFAX: 202-887-0357
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-046-578-11

Query Match      100.0%; Score 9; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
   |||||||
DB 13 TAATATTAC 5

RESULT 10
PCT-US91-05802-10
; Sequence 10, Application PC/TUS9105802
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn C.
; APPLICANT: Ecker, David J.
; TITLE OF INVENTION: Inhibition of Candida
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place-- 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb
; MEDIUM TYPE: STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
```

```

; SOFTWARE: WORDPERFECT 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05802
; FILING DATE: 19910815
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Licata, Jane M.
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISIS-0432
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; HYPOTHEICAL: NO
; ANTI-SENSE: YES
PCT-US91-05802-10

```

```

Query Match      100.0%; Score 9; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
Db 6 TAATATTAC 14

```

```

RESULT 11
US-08-793-701-45/c
; Sequence 45, Application US/08793701
; Patent No. 6248581
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, Brigitte
; APPLICANT: LIM, Eng Mong
; APPLICANT: PORTNOI, Denis
; APPLICANT: BERTHET, Francois-Xavier
; APPLICANT: TIMM, Juliano
; TITLE OF INVENTION: MYCOBACTERIA FUNCTIONAL SCREENING AND/OR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: c/o FINNEGAN, HENDERSON, FARRABOW, GARRETT &
; ADDRESSEE: DUNNER, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,701
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR9501133
; FILING DATE: 30-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94/10585
; FILING DATE: 02-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonell, Leslie A.
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02356.0075

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4132
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
US-08-793-701-45

```

```

Query Match      100.0%; Score 9; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
Db 16 TAATATTAC 8

```

```

RESULT 12
US-09-051-914-10
; Sequence 10, Application US/09051914A
; Patent No. 6096306
; GENERAL INFORMATION:
; APPLICANT: BRAVO, Alejandra
; APPLICANT: LERECCLUS, Didier
; APPLICANT: SALAMITOU, Sylvie
; APPLICANT: SANCHIS, Vincent
; APPLICANT: AGAISSE, Herve
; TITLE OF INVENTION: NEW STRAINS OF BACILLUS THURINGIENSIS AND PESTICIDE
; FILE REFERENCE: 0660-0131-0X PCT
; CURRENT APPLICATION NUMBER: US/09/051,914A
; CURRENT FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: PCT/FR96/01684
; EARLIER FILING DATE: 1996-10-28
; EARLIER APPLICATION NUMBER: FR 95/12741
; EARLIER FILING DATE: 1995-10-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 35
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
US-09-051-914-10

```

```

Query Match      100.0%; Score 9; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
Db 19 taatattac 27

```

```

RESULT 13
US-08-805-191-7
; Sequence 7, Application US/08805191
; Patent No. 5908924
; GENERAL INFORMATION:
; APPLICANT: Burdette, Douglas S
; APPLICANT: Zeikus, Joseph G
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THE GENE
; TITLE OF INVENTION: ENCODING THERMOANEROBACTER ETHANOLICUS 39E
; TITLE OF INVENTION: SECONDARY-ALCOHOL DEHYDROGENASE AND ENZYME BIOCHEMICAL
; CHARACTERIZATION
; NUMBER OF SEQUENCES: 7

```

us-09-462-955-3.rni

Thu Oct 25 13:08:47 2001

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,191
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/012,331
FILING DATE: 27-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 660336.90608
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "primer"
US-08-805-191-7

Query Match 100.0%; Score 9; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 19 TAATATTAC 27

RESULT 14
US-08-433-126A-39
Sequence 39, Application US/08433126A
Patent No. 5688935
GENERAL INFORMATION:
APPLICANT: STEPHENS, ANDREW
APPLICANT: SCHNEIDER, DAN
TITLE OF INVENTION: NUCLEIC ACID LIGANDS OF TISSUE
TITLE OF INVENTION: TARGET
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM pc compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,126A
FILING DATE: 03 MAY 1995
CLASSIFICATION: 514

Query Match 100.0%; Score 9; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 43 TAATATTAC 51

RESULT 15
US-08-433-124A-39
Sequence 39, Application US/08433124A
Patent No. 5750342
GENERAL INFORMATION:
APPLICANT: STEPHENS, ANDREW
APPLICANT: SCHNEIDER, DAN
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: NUCLEIC ACID LIGANDS OF TISSUE
TITLE OF INVENTION: TARGET
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM pc compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,124A
FILING DATE: 03 MAY 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215

```

```

; REFERENCE/DOCKET NUMBER: NEX31.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-433-124A-39

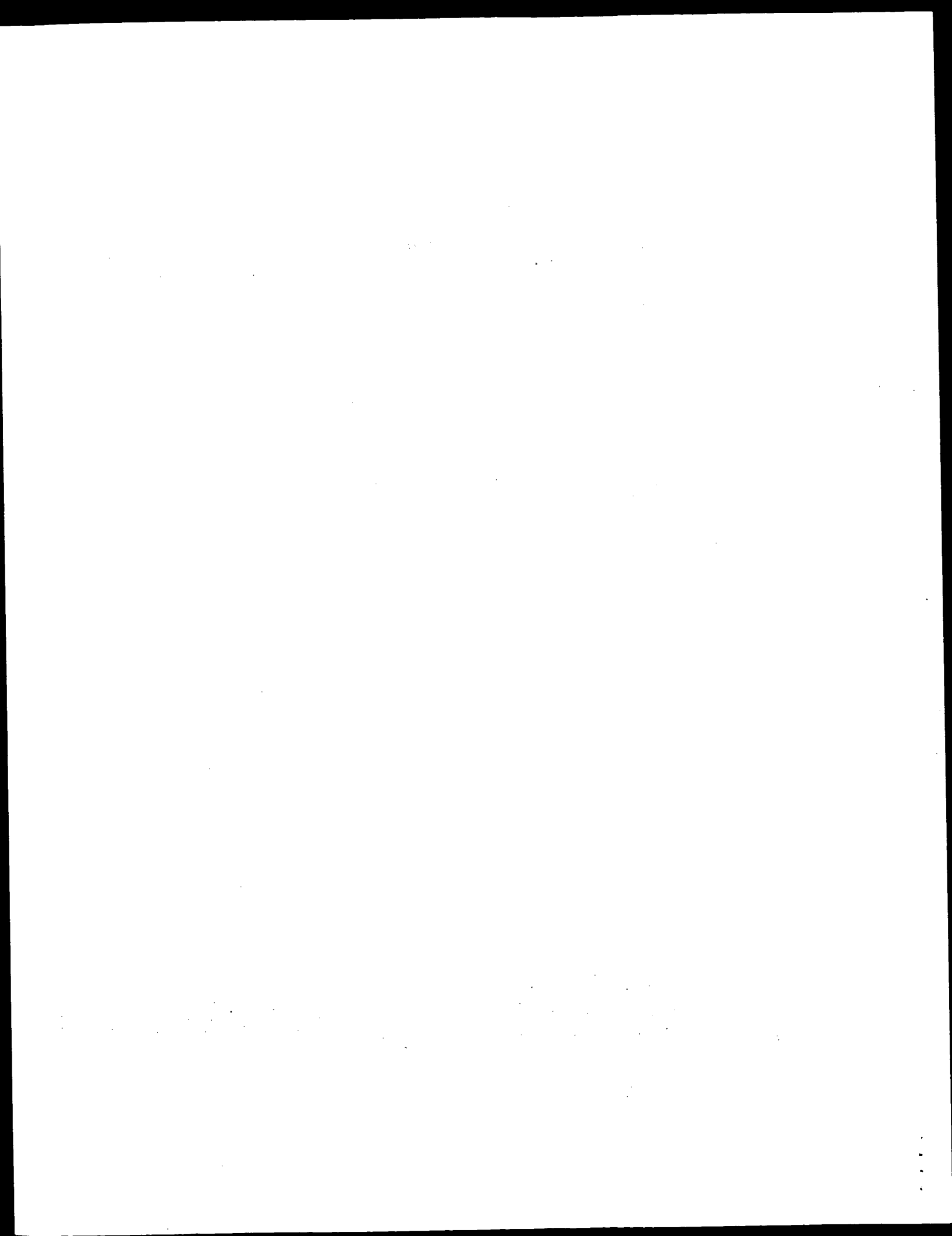
```

```

Query Match      100.0%; Score 9; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 taatattac 9
Db 43 TAATATTAC 51

```

Search completed: October 24, 2001, 10:00:31  
Job time: 216 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:27 ; Search time 393.52 seconds  
(without alignments)  
14.360 Million cell updates/sec

Title: US-09-462-955-4

Perfect score: 9

Sequence: 1 taactag 9

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description           |
|------------|-------|-------------|--------|----------|-----------------------|
| C 1        | 9     | 100.0       | 18     | AAV33149 | Plasmodium falciparum |
| C 2        | 9     | 100.0       | 18     | AAV33146 | Plasmodium vivax c    |
| C 3        | 9     | 100.0       | 22     | AAA30640 | Human G protein-co    |
| C 4        | 9     | 100.0       | 22     | AAZ50574 | 3' PCR primer-1 to    |
| C 5        | 9     | 100.0       | 22     | AAZ50574 | Double-stranded DN    |
| C 6        | 9     | 100.0       | 23     | AAV45715 | Human NKCC2 gene e    |
| C 7        | 9     | 100.0       | 28     | AAZ61141 | Reverse PCR primer    |
| C 8        | 9     | 100.0       | 30     | AAZ87693 | Human TPST-1 cDNA     |
| C 9        | 9     | 100.0       | 30     | AAZ99717 | PCR primer for cDN    |
| C 10       | 9     | 100.0       | 31     | AAQ56057 | Coconut Foliar Dec    |
| C 11       | 9     | 100.0       | 31     | AAZ02688 | Gemini virus DNA f    |

|      |   |       |     |    |           |                     |
|------|---|-------|-----|----|-----------|---------------------|
| 12   | 9 | 100.0 | 31  | 20 | AAZ02686  | Gemini virus DNA f  |
| C 13 | 9 | 100.0 | 32  | 18 | AAZ90052  | Primer for chicken  |
| C 14 | 9 | 100.0 | 35  | 18 | AAZ77100  | Pisum sativum plas  |
| C 15 | 9 | 100.0 | 35  | 21 | AAA39333  | Z. paucivorans 16S  |
| C 16 | 9 | 100.0 | 36  | 20 | AAZ28348  | PCR primer PC641 f  |
| C 17 | 9 | 100.0 | 38  | 18 | AAZ62191  | Primer PF668 for C  |
| C 18 | 9 | 100.0 | 38  | 20 | AAZ28335  | PCR primer PF668 f  |
| C 19 | 9 | 100.0 | 38  | 21 | AAZ61496  | Primer PCR2AS for   |
| C 20 | 9 | 100.0 | 38  | 22 | AAZ91511  | CmR gene/PorA prom  |
| C 21 | 9 | 100.0 | 38  | 22 | AAZ91519  | Human G protein-co  |
| C 22 | 9 | 100.0 | 39  | 21 | AAA30642  | 3' PCR primer-2 to  |
| C 23 | 9 | 100.0 | 39  | 21 | AAZ50576  | Linker coding for   |
| C 24 | 9 | 100.0 | 39  | 22 | AAZ91503  | CmR gene/PorA prom  |
| C 25 | 9 | 100.0 | 41  | 15 | AAZ97998  | Linker coding for   |
| C 26 | 9 | 100.0 | 53  | 21 | AAA48234  | T. reesei xylanase  |
| C 27 | 9 | 100.0 | 62  | 15 | AAZ069204 | Variable oligonucle |
| C 28 | 9 | 100.0 | 67  | 21 | AAA48233  | Oligo encoding N-t  |
| C 29 | 9 | 100.0 | 74  | 22 | AAZ31601  | T. reesei xylanase  |
| C 30 | 9 | 100.0 | 93  | 16 | AAZ26341  | Human gene signatu  |
| C 31 | 9 | 100.0 | 98  | 21 | AAZ09195  | Human secreted pro  |
| C 32 | 9 | 100.0 | 101 | 16 | AAZ97387  | Human type i stero  |
| C 33 | 9 | 100.0 | 110 | 18 | AAZ75998  | Staphylococcus aur  |
| C 34 | 9 | 100.0 | 112 | 21 | AAZ17499  | Human secreted pro  |
| C 35 | 9 | 100.0 | 114 | 21 | AAZ94806  | Soybean microsate   |
| C 36 | 9 | 100.0 | 116 | 21 | AAZ10236  | Human secreted pro  |
| C 37 | 9 | 100.0 | 127 | 18 | AAZ76841  | Staphylococcus aur  |
| C 38 | 9 | 100.0 | 138 | 21 | AAZ18454  | Human secreted pro  |
| C 39 | 9 | 100.0 | 138 | 21 | AAZ20546  | Human secreted pro  |
| C 40 | 9 | 100.0 | 149 | 21 | AAZ45738  | Human secreted exp  |
| C 41 | 9 | 100.0 | 151 | 21 | AAZ26107  | Human secreted pro  |
| C 42 | 9 | 100.0 | 152 | 16 | AAZ23458  | Human gene signatu  |
| C 43 | 9 | 100.0 | 159 | 21 | AAZ45500  | Human secreted exp  |
| C 44 | 9 | 100.0 | 165 | 21 | AAZ22885  | Human secreted pro  |
| C 45 | 9 | 100.0 | 177 | 19 | AAZ10779  | Human biallelic po  |

# ALIGNMENTS

|          |   |
|----------|---|
| RESULT 1 | AAV33149/c  |
| ID       | AAV33149 standard; DNA; 18 BP.                                    |
| XX       | XX  |
| AC       | AAV33149;   |
| XX       | XX  |
| DT       | 07-DEC-1998 (first entry)   |
| XX       | XX  |
| DE       | Plasmodium falciparum coxI gene PCR primer PFL.                   |
| XX       | XX  |
| KW       | Malaria; infection; therapy; diagnosis; mitochondrion; coxI gene; |
| KW       | cytochrome C oxidase; PCR; primer; ss.                            |
| XX       | XX  |
| OS       | Synthetic.  |
| OS       | Plasmodium falciparum.  |
| XX       | XX  |
| PN       | WO9835057-A1.   |
| XX       | XX  |
| PD       | 13-AUG-1998.  |
| XX       | XX  |
| PF       | 05-FEB-1998; 98WO-IB00212.  |
| XX       | XX  |
| PR       | 26-SEP-1997; 97AU-0009481.  |
| PR       | 06-FEB-1997; 97AU-0004953.  |
| XX       | XX  |
| PA       | 21-APR-1997; 97AU-0006329.  |
| XX       | XX  |
| PA       | (MOLE-) INST MOLECULAR & CELL BIOLOGY.                            |
| XX       | XX  |
| XX       | (UYSI-) UNIV SINGAPORE NAT.                                       |
| PI       | Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;                  |
| XX       | XX  |
| DR       | WPI; 1998-447251/38.  |
| XX       | XX  |
| PI       | Detecting Plasmodium infection from hybridisation with            |

PT extrachromosomal element - providing genus or species specific  
 PT diagnosis with few false negatives, in humans or animals  
 XX  
 PS Claim 6; Page 66; 120pp; English.  
 XX  
 CC This synthetic oligonucleotide sequence (PFI) is derived from the  
 CC Plasmodium vivax mitochondrial cytochrome C oxidase (coxI)  
 CC gene. In a claimed method, Plasmodium is detected in a human or  
 CC animal sample (especially blood or dried blood) by treating it, or  
 CC element of derived nucleic acid (A) and detecting any hybridisation.  
 CC (A) can include the PSI-PL470, PLH-PPH, PRB or PWQ gene (see  
 CC AAV33135-38), the coxI gene or nucleic acids derived from them. Also  
 CC new are (A)-specific probes and primers (see AAV33139-56). The method  
 CC is used to diagnose Plasmodium infection. The high degree of  
 CC similarity between (A) from different species allows development of  
 CC genus- or species-specific assays that result in fewer false  
 CC negatives than known methods (typically 1% against 3%). Primer PFI  
 CC can be used with another primer (see AAV33145) for the species-  
 CC specific detection of P. falciparum.  
 XX  
 SQ Sequence 18 BP; 4 A; 3 C; 3 G; 8 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 Db 10 TAATACTAG 2

RESULT 2  
 ID AAV33146 standard; DNA; 18 BP.  
 XX  
 AC AAV33146;  
 XX  
 DT 07-DEC-1998 (first entry)  
 XX  
 DE Plasmodium vivax coxI gene PCR primer PV1.  
 XX  
 KW Malaria; infection; therapy; diagnosis; mitochondrion; coxI gene;  
 KW cytochrome C oxidase; PCR; primer; ss.  
 XX  
 OS Synthetic.  
 OS Plasmodium vivax.  
 XX  
 PN WO9835057-A1.  
 XX  
 PD 13-AUG-1998.  
 XX  
 PF 05-FEB-1998; 98WO-IB00212.  
 XX  
 PR 26-SEP-1997; 97AU-0009481.  
 PR 06-FEB-1997; 97AU-0004953.  
 PR 21-APR-1997; 97AU-0006329.  
 XX  
 XX (MOLE-) INST MOLECULAR & CELL BIOLOGY.  
 PA (UYSI-) UNIV SINGAPORE NAT.  
 XX  
 XX Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;  
 XX WPI; 1998-447251/38.  
 DR  
 XX  
 XX Detecting Plasmodium infection from hybridisation with  
 XX extrachromosomal element - providing genus or species specific  
 PT diagnosis with few false negatives, in humans or animals  
 PT  
 XX  
 XX Claim 6; Page 65; 120pp; English.  
 PS  
 XX This synthetic oligonucleotide sequence (PFI) is derived from the

CC Plasmodium vivax mitochondrial cytochrome C oxidase (coxI) gene.  
 CC In a claimed method, Plasmodium is detected in a human or animal  
 CC sample by treating it, or derived nucleic acid, with a plasmodium  
 CC extrachromosomal genetic element or derived nucleic acid (A) and  
 CC detecting any hybridisation. (A) can include the PSI-PL470,  
 CC PLH-PPH, PRB or PWQ gene (see AAV33135-38), the coxI gene or nucleic  
 CC acids derived from them. Also new are (A)-specific probes and  
 CC primers (see AAV33139-56). The method is used to diagnose Plasmodium  
 CC infection. The sample can be blood or dried blood. The high  
 CC degree of similarity between (A) from different species allows  
 CC development of genus- or species-specific assays that result in  
 CC fewer false negatives than known methods (typically 1% against 3%).  
 CC Primer PFI can be used with another primer (see AAV33145) for the  
 CC species-specific detection of P. vivax.

XX Sequence 18 BP; 3 A; 3 C; 3 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 Db 10 TAATACTAG 2

RESULT 3  
 ID AAA30640 standard; DNA; 22 BP.  
 XX  
 AC AAA30640;  
 XX  
 DT 21-AUG-2000 (first entry)  
 XX  
 DE Human G protein-coupled receptor GHSR PCR primer, SEQ ID NO:84.  
 DE G protein-coupled receptor; GPCR; constitutively active;  
 KW intracellular loop 3; transmembrane domain 6; drug screening;  
 KW agonist; antagonist; PCR primer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200022129-A1.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 12-OCT-1999; 99WO-US23938.  
 XX  
 PR 13-OCT-1998; 98US-0170496.  
 XX  
 PA (AREN-) ARENA PHARM INC.  
 XX  
 PI Behan DP, Chalmers DT, Liaw CW;  
 XX WPI; 2000-329165/28.  
 DR  
 XX Non-endogenous constitutively activated human G protein-coupled  
 PT receptors, useful for identifying agonists for use as pharmaceutical  
 PT agents  
 XX  
 PS Example 1; Page 42; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions  
 CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-  
 CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743  
 CC and AAA30775-A30779). The mutant proteins of the invention contain a  
 CC mutation in a portion of the protein comprising intracellular loop 3  
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,  
 CC is substituted for an endogenous residue in IC3 at a position 16 amino  
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence  
 CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg  
 CC or Ala, and is preferably Lys. When the endogenous residue at this

position is Lys, this residue is replaced by His, Arg or preferably Ala. The 15 amino acid stretch between the substituted amino acid and the Pro may be endogenous, non-endogenous, or a mixture of endogenous and non-endogenous residues. The constitutively active GPCRs are useful for identifying antagonists, agonists and partial agonists for use as pharmaceutical agents. The mutant proteins are also useful in research settings for elucidating the roles of the receptors in normal and diseased conditions. Antagonists for a particular GPCR are useful for treating diseases and disorders associated with that receptor. Because the novel mutant GPCRs are constitutively active, they can be used directly for screening of compounds without the need for endogenous ligands. The present sequence represents a PCR primer used in an exemplification of the invention to isolate cDNA encoding a human wild-type GPCR for cloning, an exemplification of the invention. The GPCR cDNA was subjected to site-directed mutagenesis (SDM) to generate DNA encoding the corresponding mutant of the invention.

Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 taatactag 9  
Db 9 taatactag 17

RESULT 4  
AAZ50574  
ID AAZ50574 standard; DNA; 22 BP.  
XX  
AC AAZ50574;  
XX  
20-JUN-2000 (first entry)  
XX  
3' PCR primer-1 to amplify human GHSR DNA.  
XX  
G protein-coupled orphan receptor; GPCR; agonist; G protein;  
GPCR fusion protein; inverse agonist; drug; treatment; PCR primer;  
GHSR; G protein-coupled receptor; human; ss.  
XX  
Homo sapiens.  
XX  
WO200006597-A2.  
XX  
10-FEB-2000.

30-JUL-1999; 99WO-US17425.  
XX  
31-JUL-1998; 98US-0094879.  
XX  
30-OCT-1998; 98US-0106300.  
XX  
04-DEC-1998; 98US-0110906.  
XX  
26-FEB-1999; 99US-0121851.  
XX  
(AREN-) ARENA PHARM INC.  
XX  
Behan DP, Chalmers DT, Liaw C, Lin I, Lowitz K, Chen R;  
XX  
WPI; 2000-195260/17.  
XX  
Identification of a compound useful as a therapeutic agent, comprises  
PT identifying a compound against constitutively activated G  
PT protein-coupled orphan receptors -  
XX  
Example 2; Page 27; 123pp; English.  
XX

The patent discloses a method of identifying agonists and inverse or partial agonists to the endogenous, constitutively activated G protein-coupled orphan receptors (GPCRs), by contacting them with a GPCR fusion protein comprising a GPCR and a G protein. Determining expression of GPCRs in tissue samples can be used to identify related

CC diseases. Inverse agonists to these receptors can be used as drugs for  
CC treating GPCR-related diseases. The present sequence is a 3' PCR primer,  
CC used in primary PCR reaction to amplify human GHSR DNA from  
CC hippocampus cDNA template.

SQ Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 taatactag 9  
Db 9 taatactag 17

RESULT 5  
AAF31603  
ID AAF31603 standard; DNA; 22 BP.  
XX  
AC AAF31603;  
XX  
09-APR-2001 (first entry)  
XX  
Double-stranded DNA tag PCR primer pTag2.  
XX  
attenuated microorganism; signature tagged transposon mutant;  
KW mutant library; mycobacterial infection; actinomycetales;  
KW antibacterial; immunostimulant; vaccine; PCR primer; ss.  
XX  
OS Unidentified.

XX  
PN WO200102555-A1.  
XX  
11-JAN-2001.

XX  
PF 06-JUL-2000; 2000WO-IB00950.

XX  
PR 06-JUL-1999; 99US-0142982.

XX  
PR 08-JUL-1999; 99US-0142833.

XX  
(INSP) INST PASTEUR.

XX  
Gicquel B, Guilhot C, Camacho L;

XX  
WPI; 2001-091804/10.

Screening a mutant library for mutants unable to grow under specific conditions and for identifying loci involved in pathogenicity, comprises using signature tagged transposon mutagenesis -

XX  
Example 4; Page 19; 159pp; English.

The present sequence is given in a specification relating to a method for screening a library of mutants. The method comprises constructing a library with insertions in genes and/or regulatory regions of the organisms of interest, where the insertion contains a tag and/or a transposon associated with a tag. The mutants are identified by hybridisation of the tags to known sequences. The method is useful for treating an individual suffering from a mycobacterial infection, suspected of being infected with a Mycobacterium, or having been exposed to an infectious Mycobacterium. It is also useful for identifying and isolating mutants of actinomycetales and for identifying mutants of microorganisms, preferably an actinomycetales, such as M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and M. paratuberculosis, that is unable to grow under specific conditions. It is especially useful for identifying loci involved in pathogenicity. It is useful in constructing vaccines. The method can be used to screen multiple libraries concurrently. It can screen libraries of different organisms or different strains of the same organism. The present sequence was used to create the tag used in the construction of tagged

CC mutants.  
 XX Sequence 22 BP; 7 A; 5 C; 1 G; 9 T; 0 other;  
 SQ

Query Match 100.0%; Score 9; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9  
 |||||  
 Db 3 taatactag 11

RESULT 6  
 AAV45715  
 ID AAV45715 standard; DNA; 23 BP.  
 AC AAV45715;  
 XX  
 XX 21-DEC-1998 (first entry)  
 DT  
 DE Human NKCC2 gene exon 19 forward primer hNKCC2ex19.  
 XX  
 XX Na-K-2Cl cotransporter; NKCC2; human; Bartter's syndrome;  
 KW ion transport; hypokalaemic alkalosis; hypercalciuria;  
 KW nephrocalcinosis; diagnosis; therapy; SSCP; primer; ss.  
 XX  
 XX Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO9829431-A1.  
 PN  
 PD 09-JUL-1998.  
 XX  
 XX 19-DEC-1997; 97WO-US23553.  
 PF  
 XX 31-DEC-1996; 96US-0778052.  
 PR  
 XX (UYVA ) UNIV YALE.  
 PA  
 XX Lifton RP, Simon DB;  
 PI WPI; 1998-388029/33.  
 DR  
 XX Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2Cl  
 PT cotransporter genes - useful for developing products for the  
 PT diagnosis and treatment of ion transport disorders, e.g. Gitelman's  
 PT Syndrome or Bartter's Syndrome  
 XX  
 XX Example 2; Page 65; 105pp; English.

XX Primers hNKCC2ex19 forward and reverse (see AAV45715 and AAV45716,  
 CC respectively) are designed to amplify exon 19 of the human NKCC2  
 CC gene (see AAV4562) that codes for Na-K-2Cl cotransporter NKCC2  
 CC protein (see AAV23683). Both primers are located within introns of  
 CC the gene. 27 Sets of specific primers (see AAV45677-V45730) were  
 CC used for SSCP analysis of NKCC2. Amplified products were analysed  
 CC for molecular variants by electrophoresis, and identified variants  
 CC were sequenced. Complete linkage of Bartter's syndrome with NKCC2  
 CC was demonstrated. Identification of the molecular basis of  
 CC Bartter's syndrome allows for the genetic diagnosis of this  
 CC disorder. The invention provides products and methods useful for  
 CC diagnosis and treatment of Bartter's syndrome and other ion  
 CC transport disorders.  
 XX  
 XX Sequence 23 BP; 8 A; 4 C; 5 G; 6 T; 0 other;  
 SQ

Query Match 100.0%; Score 9; DB 19; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9  
 |||||  
 Db 19 TAATACTAG 11

Query Match 100.0%; Score 9; DB 21; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9  
 |||||  
 Db 19 TAATACTAG 11

QY 1 taatactag 9  
 |||||  
 Db 2 taatactag 10

RESULT 7  
 AAZ61141/C  
 ID AAZ61141 standard; DNA; 28 BP.  
 XX  
 XX AAZ61141;  
 AC  
 XX 30-MAY-2000 (first entry)  
 DT  
 DE Reverse PCR primer used to amplify DNA encoding the Shiga toxin.  
 XX  
 XX Chemokine receptor; ligand; inflammatory response; immune effector cell;  
 KW secondary tissue damage; central nervous system injury; Shiga;  
 KW CNS inflammatory disease; neurodegenerative disorder; heart disease;  
 KW inflammatory eye disease; inflammatory bowel disease; PCR primer;  
 KW inflammatory joint disease; inflammatory kidney; renal disease;  
 KW inflammatory lung disease; inflammatory nasal disease; thyroiditis;  
 KW inflammatory thyroid disease; cytokine-regulated cancer; ss.  
 XX  
 XX Shigella dysenteriae.  
 OS  
 XX WO200004926-A2.  
 PN  
 XX 03-FEB-2000.  
 PD  
 XX 21-JUL-1999; 99WO-CA00659.  
 PF  
 XX 22-JUL-1998; 98US-0120523.  
 PR  
 XX (OSPR-) OSPREY PHARM LTD.  
 PA  
 XX McDonald JR, Coggins PJ;  
 PI WPI; 2000-182542/16.  
 DR  
 XX A new therapeutic agent comprising a conjugate for treating secondary  
 PT tissue damage and other disease conditions like Alzheimer's disease,  
 PT stroke, Parkinson's disease and atherosclerosis  
 XX  
 XX Example 1; Page 135; 204pp; English.

XX PCR primers AAZ61140-42 were used to amplify nucleic acids encoding  
 CC the Shiga toxin. The toxin can be incorporated into the conjugates  
 CC of the invention. The specification describes a conjugate, comprising  
 CC a targeted agent and a chemokine receptor ligand. The conjugate  
 CC binds to a chemokine receptor resulting in internalisation of the  
 CC targeted agent in cells bearing the receptor. The conjugates are  
 CC used for formulating a medicament or for treating disorders associated  
 CC with inflammatory responses resulting from activation, proliferation  
 CC and migration of immune effector cells. The disorders or disease states  
 CC comprise secondary tissue damage such as central nervous system (CNS)  
 CC injury, CNS inflammatory diseases, neurodegenerative disorders, heart  
 CC disease, inflammatory eye diseases, inflammatory bowel diseases,  
 CC inflammatory joint diseases, inflammatory kidney or renal diseases,  
 CC inflammatory lung diseases, inflammatory nasal diseases, inflammatory  
 CC thyroid disease such as thyroiditis, or cytokine-regulated cancers.  
 XX  
 XX Sequence 28 BP; 6 A; 7 C; 8 G; 7 T; 0 other;  
 SQ

RESULT 8  
 AAX87693/c  
 ID AAX87693 standard; DNA; 30 BP.  
 XX  
 AC AAX87693;  
 XX  
 DT 26-OCT-1999 (first entry)  
 XX  
 DE Human TPST-1 cDNA bottom strand PCR primer.  
 XX  
 DE TPST-1; tyrosylprotein sulfotransferase; tyrosine O-sulfation;  
 KW post-translation modification; human; PCR; primer; ss.  
 KW  
 XX  
 OS Synthetic.  
 OS  
 OS Homo sapiens.  
 XX  
 PN WO9938980-A2.  
 XX  
 PD 05-AUG-1999.  
 XX  
 PF 28-JAN-1999; 99WO-US01756.  
 XX  
 PR 09-SEP-1998; 98US-0150141.  
 PR  
 PR 29-JAN-1998; 98US-0072994.  
 PR  
 PR 09-SEP-1998; 98US-0150133.  
 XX  
 PA (OKLA ) UNIV OKLAHOMA STATE.  
 XX  
 PI Moore KL;  
 XX  
 DR WPI; 1999-494094/41.  
 XX  
 PT Human, murine and Caenorhabditis elegans tyrosylprotein  
 PT sulfotransferases, useful for post-translational tyrosine sulfation  
 XX  
 PS Example; Page 39; 123pp; English.  
 XX  
 CC This bottom strand primer was used with a top strand primer (see  
 CC AAX87692) in the PCR amplification of the human tyrosylprotein  
 CC sulfotransferase (TPST-1) coding sequence (see AAX87686) using EST  
 CC clone #116978 as template. The primers introduce a 5' BamHI site  
 CC and a 3' SpeI site. The amplified cDNA was ligated into vector  
 CC pCDNA3.1(+), and TPST-1 (see AAY06623) was expressed as a fusion  
 CC protein containing an N-terminal epitope for HPCA (see AAY06630)  
 CC in mammalian cells. TPST-1 is useful for the post-translational  
 CC tyrosine O-sulfation of proteins and peptides.  
 XX  
 SQ Sequence 30 BP; 5 A; 9 C; 6 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 taactactag 9  
 |||||  
 Db 12 TAATACTAG 4  
 RESULT 9  
 AAZ99717/c  
 ID AAZ99717 standard; DNA; 30 BP.  
 XX  
 AC AAZ99717;  
 XX  
 DT 12-JUL-2000 (first entry)  
 XX  
 DE PCR primer for cDNA encoding human tyrosylprotein sulfotransferase 1.  
 XX  
 DE Human; tyrosylprotein sulfotransferase 1; TPST-1; tyrosine O-sulfation;  
 KW P-selectin glycoprotein ligand-1; PSGL-1; anti-inflammatory;  
 KW neutrophil binding; PCR primer; ss.  
 KW  
 XX

Query Match 100.0%; Score 9; DB 20; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 taactactag 9  
 |||||  
 Db 12 TAATACTAG 4

RESULT 9  
 AAZ99717/c  
 ID AAZ99717 standard; DNA; 30 BP.  
 XX  
 AC AAZ99717;  
 XX  
 DT 12-JUL-2000 (first entry)  
 XX  
 DE PCR primer for cDNA encoding human tyrosylprotein sulfotransferase 1.  
 XX  
 DE Human; tyrosylprotein sulfotransferase 1; TPST-1; tyrosine O-sulfation;  
 KW P-selectin glycoprotein ligand-1; PSGL-1; anti-inflammatory;  
 KW neutrophil binding; PCR primer; ss.  
 KW  
 XX

OS Homo sapiens.  
 XX  
 PN WO200014250-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 23-JUL-1999; 99WO-US16750.  
 XX  
 PR 09-SEP-1998; 98US-0150133.  
 PR  
 PR 09-SEP-1998; 98US-0150141.  
 PR  
 PR 28-JAN-1999; 99WO-US01756.  
 XX  
 PA (OKLA ) UNIV OKLAHOMA STATE.  
 XX  
 PI Moore KL;  
 XX  
 DR WPI; 2000-271057/23.  
 XX  
 PT New human or murine tyrosylprotein sulfotransferase, used for in vitro  
 PT or in vivo sulfation of proteins and to screen for anti-inflammatory  
 PT agents -  
 XX  
 PS Example; Page 40; 141pp; English.  
 XX  
 CC PCR primers AAZ99716-17 were used to amplify cDNA encoding human  
 CC tyrosylprotein sulfotransferase 1 (TPST-1) polypeptide. TPST-1 catalyses  
 CC the transfer of sulfate from 3'-phosphoadenosine-5'-phosphosulfate to  
 CC tyrosine residues (tyrosine O-sulfation) in proteins. TPST-1  
 CC polypeptides are used for in vitro or in vivo sulfation of proteins or  
 CC peptides, particularly P-selectin glycoprotein ligand-1 (PSGL-1) or its  
 CC functional fragments. They are also used to raise specific antibodies,  
 CC and are useful as immunoassay reagents, and to identify specific  
 CC inhibitors. These inhibitors are potentially useful as anti-inflammatory  
 CC agents, by inhibiting PSGL-1 mediated binding of neutrophils to  
 CC endothelial cells. The TPST-1 polynucleotide is useful for identifying  
 CC polymorphisms that may be associated with a disease phenotype.  
 XX  
 SQ Sequence 30 BP; 5 A; 9 C; 6 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 taactactag 9  
 |||||  
 Db 12 TAATACTAG 4

RESULT 10  
 AAZ99717/c  
 ID AAZ99717 standard; DNA; 31 BP.  
 XX  
 AC AAZ99717;  
 XX  
 DT 12-AUG-1994 (first entry)  
 XX  
 DE Coconut Foliar Decay Virus promoter stem-loop.  
 XX  
 DE Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;  
 KW phloem-specific; stem-loop structure; transgenic plant; ds.  
 KW  
 XX  
 OS Coconut Foliar Decay Virus.  
 XX  
 PI Key Location/Qualifiers  
 stem\_loop 1..31  
 /tag= a  
 /function= promoter  
 /note= "loop has homology to geminivirus sequence"  
 FT DE4306832-C.  
 PN  
 XX  
 XX 24-FEB-1994.

Thu Oct 25 13:08:51 2001

```

XX PF 04-MAR-1993; 93DE-4306832.
XX PF 04-MAR-1993; 93DE-4306832.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1994-058406/08.
XX DR Use of coconut foliar decay virus DNA as promoter - for
XX PT tissue-specific gene expression in transgenic plants
XX PT Claim 1; Fig 2; 8pp; German.
XX PS A DNA fragment from the CFV genome can be used as a phloem-specific
XX CC promoter in the construction of transgenic plants. The promoter is
XX CC strong; it has 30-50% of the activity of the CaMV 35S promoter in
XX CC tobacco plants.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
Db 12 taatactag 20

RESULT 11
AA02688
ID AAX02688 standard; DNA; 31 BP.
XX AC AAX02688;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX OS Gemini virus.
XX FH Key Location/Qualifiers
XX FT stem_loop 1..31 /*tag= a
XX FT misc_binding 1..11 /*tag= b
XX FT /*note= "Region binds to nucleotides 21 to 31"
XX FT misc_binding 21..31 /*tag= c
XX FT /*note= "Region binds to nucleotides 1 to 11"
XX PN DE19730502-A1.
XX PD 21-JAN-1999.
XX PF 16-JUL-1997; 97DE-1030502.
XX PR 16-JUL-1997; 97DE-1030502.
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX XX WPI; 1999-096863/09.
XX XX Coconut foliar decay virus promoters - for gene expression in

```

```

PT bacteria and yeasts
XX Disclosure; Fig 2; 14pp; German.
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX that includes the stem-loop structure of CFV DNA but lacks the
XX translation start codons of open reading frames ORF1 and/or ORF2. The
XX new CFV DNA fragment is useful as a bacterial or yeast promoter, as a
XX promoter for tissue-specific (especially phloem-specific) gene expression
XX in plants and for production of chimeric constructs for transient or
XX stable expression. Certain fragments of CFV DNA have stronger promoter
XX activity in E. coli than the CaMV 35S promoter.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
Db 12 taatactag 20

RESULT 12
AA02686
ID AAX02686 standard; DNA; 31 BP.
XX AC AAX02686;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX OS Gemini virus.
XX FH Key Location/Qualifiers
XX FT stem_loop 1..31 /*tag= a
XX FT misc_binding 1..11 /*tag= b
XX FT /*note= "Region binds to nucleotides 21 to 31"
XX FT misc_binding 21..31 /*tag= c
XX FT /*note= "Region binds to nucleotides 1 to 11"
XX PN DE19730535-A1.
XX PD 21-JAN-1999.
XX PF 16-JUL-1997; 97DE-1030535.
XX PR 16-JUL-1997; 97DE-1030535.
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX WPI; 1999-096867/09.
XX Coconut foliar decay virus promoters - for gene expression in
XX plants, bacteria and yeasts
XX Disclosure; Fig 2; 12pp; German.
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX that includes the stem-loop structure of CFV DNA but lacks the
XX translation start codons of open reading frames ORF1 and/or ORF2. The
XX new CFV DNA fragment is useful as a bacterial or yeast promoter, as a

```

CC promoter for tissue-specific (especially phloem-specific) gene expression  
 CC in plants and for production of chimeric constructs for transient or  
 CC stable expression. Certain fragments of CPDV DNA have stronger promoter  
 CC activity in *E. coli* than the CaMV 35S promoter.  
 XX  
 SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 taatactag 9  
 |||||  
 Db 12 taatactag 20

RESULT 13  
 AAT90052/c  
 ID AAT90052 standard; DNA; 32 BP.  
 AC AAT90052;  
 XX  
 DT 27-MAR-1998 (first entry)  
 XX  
 DE Primer for chicken urokinase cDNA.  
 XX  
 KW Low molecular weight; chicken; urokinase; reporter; screening;  
 KW PCR primer; ss.  
 XX  
 OS Gallus domesticus.  
 XX  
 PN WC9734144-A1.  
 XX  
 PD 18-SEP-1997.  
 XX  
 PF 11-MAR-1997; 97WO-EP01219.  
 XX  
 PR 20-DEC-1996; 96DE-4054616.  
 PR 14-MAR-1996; 96DE-4011780.  
 XX  
 PA (SCHD ) SCHERING AG.  
 XX  
 PI Langer G, Schleuning W, Toschi L;  
 XX  
 DR WPI; 1997-470977/43.  
 XX  
 PT Determining effect of test substance on cultured cells - by  
 PT transfecting cells with plasmid having gene for low molecular weight  
 PT hen urokinase as reporter, which can be detected by plasminogen  
 PT activation and colour forming or fluorochrome substrate reaction  
 XX  
 PS Example 1; Page 16; 32pp; German.

CC The present sequence was used in the development of a method for  
 CC determining the effect of test substance in a cell culture system.  
 CC The method comprises preparing a plasmid having an expression  
 CC cassette including the low molecular weight gene for chicken  
 CC urokinase as a reporter, transfecting eukaryotic cells  
 CC with the plasmid, inducing the cells with the test substance,  
 CC harvesting the cell supernatant and detecting urokinase by adding  
 CC chicken plasminogen (this addition precisely defines the start of  
 CC the reaction), so that this is converted to plasmin by the  
 CC urokinase, and detecting the plasmin in the supernatant with a  
 CC highly sensitive chromogenic or fluorochromic substrate. The method  
 CC can be used to screen new compounds, non-steroidal glucocorticoids  
 CC or anti-glucocorticoids, determine receptors present in cells,  
 CC particularly steroid hormone receptors and investigate promoter  
 CC dependent modulated transcription processes. The method does not  
 CC require radioisotopes or extraction of reporter gene products, is  
 CC sensitive, non-destructive and can be performed in microtitre  
 CC plates. Urokinase is small (so easily separated from endogenous  
 CC proteases by gel electrophoresis), and unlike other reporter gene

CC products is stable in culture supernatant, permitting direct  
 CC measurement. The method can be calibrated using a precise number of  
 CC transfected cells, avoiding difficulties of calibration based on  
 CC enzymatic activity. The urokinase/plasminogen reaction represents a  
 CC long lived amplification system, which is well characterised.  
 XX  
 SQ Sequence 32 BP; 5 A; 7 C; 11 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 18; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 taatactag 9  
 |||||  
 Db 20 TAATACTAG 12

RESULT 14  
 AAT77100/c  
 ID AAT77100 standard; DNA; 35 BP.  
 XX  
 AC AAT77100;  
 XX  
 DT 20-JAN-1998 (first entry)  
 XX  
 DE Pisum sativum plastocyanin promoter sub-sequence.  
 XX  
 KW plastocyanin; promoter; enhancer; reporter gene; pea;  
 KW transgenic plant; ds.  
 XX  
 OS Pisum sativum.  
 XX  
 FH Key Location/Qualifiers  
 FT enhancer 1..31  
 FT /tag= a  
 FT misc\_feature 1..4  
 FT /tag= b  
 FT /label= sticky\_end  
 FT /note= "The 5' end of complementary strand  
 FT overhangs the 3' end of this strand by the  
 FT sequence 5'-GATC-3'."  
 XX  
 PN WO9720056-A2.  
 XX  
 PD 05-JUN-1997.  
 XX  
 PF 26-NOV-1996; 96WO-GB02910.  
 XX  
 PR 29-NOV-1995; 95GB-0024350.  
 XX  
 PA (ADTE-) ADVANCED TECHNOLOGIES CAMBRIDGE.  
 XX  
 PI Gray JC, Sandhu JS, Webster CI;  
 XX  
 DR WPI; 1997-310610/28.  
 XX  
 PT Enhancer from pea plastocyanin promoter region - used to produce  
 PT transgenic plants having increased expression of a desired gene  
 XX  
 PS Claim 11; Figure 7; 46pp; English.

CC This sequence represents the pea plastocyanin promoter enhancer  
 CC which is a sub-sequence of AAT77099. This enhancer has been used to  
 CC increase the expression of the GUS reporter gene in tobacco where the  
 CC GUS gene was under the control of the PetE promoter and 35S CaMV  
 CC promoter. The enhancer was also used joined to the -330 to +1 region of  
 CC the patatin promoter pS20 which was joined to the GUS reporter and  
 CC transformed into potato using A tumefaciens. Expression of the gene  
 CC promoter of one or more genes of a plant increases in green or non-green  
 CC tissues, especially roots, tubers, seeds, flowers or leaves and  
 CC transformed plants may be dicotyledonous or monocotyledonous. This  
 CC enhancer may operate both in normal and reverse directions and can be

Thu Oct 25 13:08:51 2001

us-09-462-955-4.rng

CC attached to the promoter or terminator of the gene to be expressed.

XX  
SQ Sequence 35 BP; 17 A; 4 C; 2 G; 12 T; 0 other;

Query Match 100.0%; Score 9; DB 18; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
| | | | |  
Db 19 TAATACTAG 11

## RESULT 15

AAA39333/C  
ID AAA39333 standard; DNA; 35 BP.

XX  
AC AAA39333;

XX  
DT 11-SEP-2000 (first entry)

XX  
DE Z. paucivorans 16S rRNA and 23S rRNA gene spacer region SEQ ID NO:8.

XX  
KW 16S rRNA; 23S rRNA; spacer region; detection; microbe; beer;  
brewery; ds.

XX  
OS Zymophilus paucivorans.

XX  
PN JP2000106881-A.

XX  
PD 18-APR-2000.

XX  
PF 08-OCT-1998; 98JP-0286697.

XX  
PR 08-OCT-1998; 98JP-0286697.

XX  
PA (ASAK ) ASAHI BREWERIES LTD.

XX  
DR WPI; 2000-353477/31.

XX  
PT A gene sequence spacer region between a gene encoding 16S rRNA and a  
gene encoding 23S rRNA, useful for detection of microbes -

XX  
PS Claim 8; Page 2; 18pp; Japanese.

XX  
CC The present invention describes gene sequences of spacer region between  
a gene encoding 16S rRNA and a gene encoding 23S rRNA for the detection  
of microbes. Also described is a method for the detection of microbes  
in which an oligonucleotide prepared from a spacer region as described  
above is made to function as a primer for nucleic acid synthesis and  
treated by gene amplification. The method is used for the detection of  
microbes found in beer breweries. The present sequence represents a  
region DNA sequence, from the present invention.

XX  
SQ Sequence 35 BP; 11 A; 3 C; 7 G; 14 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
| | | | |  
Db 15 TAATACTAG 7

Search completed: October 24, 2001, 10:07:28  
Job time: 633 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:56 ; Search time 5479.82 Seconds  
(without alignments)  
15.525 Million cell updates/sec

Title: US-09-462-955-4  
Perfect score: 9  
Sequence: 1 taactagtag 9

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estin6:\*  
69: em\_estin7:\*  
70: em\_estin8:\*  
71: em\_estin9:\*  
72: em\_estin10:\*  
73: em\_estin11:\*  
74: em\_estin12:\*  
75: em\_estin13:\*  
76: em\_estin14:\*  
77: em\_estin15:\*  
78: em\_estin16:\*  
79: em\_estin17:\*  
80: em\_estin18:\*  
81: em\_estin19:\*  
82: em\_estin20:\*  
83: em\_estin21:\*  
84: em\_estin22:\*  
85: em\_estin23:\*  
86: em\_estin24:\*  
87: em\_estin25:\*  
88: em\_estin26:\*  
89: em\_estin27:\*  
90: em\_estin28:\*  
91: em\_estin29:\*  
92: em\_estin30:\*  
93: em\_estin31:\*  
94: em\_estin32:\*  
95: em\_estin33:\*  
96: em\_estin34:\*  
97: em\_estin35:\*  
98: em\_estin36:\*  
99: em\_estin37:\*  
100: em\_estin38:\*  
101: em\_estin39:\*  
102: em\_estin40:\*  
103: em\_estin41:\*  
104: em\_estin42:\*  
105: em\_estin43:\*  
106: em\_estin44:\*  
107: em\_estin45:\*  
108: em\_estin46:\*  
109: em\_estin47:\*  
110: em\_estin48:\*  
111: em\_estin49:\*  
112: em\_estin50:\*  
113: em\_estin51:\*  
114: em\_estin52:\*  
115: em\_estin53:\*  
116: em\_estin54:\*

Thu Oct 25 13:08:55 2001

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1        | 9     | 100.0       | 40     | 167   | BE386120    |
| C 2        | 9     | 100.0       | 55     | 249   | AZ769245    |
| C 3        | 9     | 100.0       | 57     | 14    | AF097417    |
| C 4        | 9     | 100.0       | 57     | 243   | AZ434302    |
| C 5        | 9     | 100.0       | 58     | 10    | AA637860    |
| C 6        | 9     | 100.0       | 59     | 113   | AW215812    |
| C 7        | 9     | 100.0       | 59     | 113   | AW215820    |
| C 8        | 9     | 100.0       | 62     | 258   | TA82F10Q    |
| C 9        | 9     | 100.0       | 66     | 250   | AZ808515    |
| C 10       | 9     | 100.0       | 67     | 24    | AA761414    |
| C 11       | 9     | 100.0       | 71     | 4     | AA250291    |
| C 12       | 9     | 100.0       | 71     | 123   | AW989354    |
| C 13       | 9     | 100.0       | 71     | 138   | BE627067    |
| C 14       | 9     | 100.0       | 71     | 138   | BE630111    |
| C 15       | 9     | 100.0       | 71     | 138   | BE688187    |
| C 16       | 9     | 100.0       | 71     | 138   | BE692587    |
| C 17       | 9     | 100.0       | 71     | 141   | BE849664    |
| C 18       | 9     | 100.0       | 73     | 167   | BE448984    |
| C 19       | 9     | 100.0       | 75     | 2     | AA079298    |
| C 20       | 9     | 100.0       | 78     | 113   | AW226900    |
| C 21       | 9     | 100.0       | 80     | 138   | BE627009    |
| C 22       | 9     | 100.0       | 84     | 2     | AA079297    |
| C 23       | 9     | 100.0       | 84     | 106   | AA012912    |
| C 24       | 9     | 100.0       | 85     | 13    | AA937702    |
| C 25       | 9     | 100.0       | 87     | 11    | AA760419    |
| C 26       | 9     | 100.0       | 87     | 249   | AZ783472    |
| C 27       | 9     | 100.0       | 90     | 155   | C01576      |
| C 28       | 9     | 100.0       | 93     | 159   | N28044      |
| C 29       | 9     | 100.0       | 94     | 20    | AI465178    |
| C 30       | 9     | 100.0       | 94     | 241   | AZ309606    |
| C 31       | 9     | 100.0       | 95     | 247   | AZ660200    |
| C 32       | 9     | 100.0       | 95     | 31    | AV562434    |
| C 33       | 9     | 100.0       | 97     | 7     | AA419984    |
| C 34       | 9     | 100.0       | 99     | 107   | AU014567    |
| C 35       | 9     | 100.0       | 100    | 10    | AA689650    |
| C 36       | 9     | 100.0       | 100    | 24    | AI718783    |
| C 37       | 9     | 100.0       | 101    | 164   | BE152001    |
| C 38       | 9     | 100.0       | 101    | 164   | BE179567    |
| C 39       | 9     | 100.0       | 102    | 166   | BE365112    |
| C 40       | 9     | 100.0       | 102    | 189   | T61718      |
| C 41       | 9     | 100.0       | 102    | 237   | AZ059959    |
| C 42       | 9     | 100.0       | 104    | 8     | AA518125    |
| C 43       | 9     | 100.0       | 104    | 121   | AW810470    |
| C 44       | 9     | 100.0       | 104    | 164   | BE145670    |
| C 45       | 9     | 100.0       | 105    | 1     | AA014683    |

ALIGNMENTS

|                  |  |       |      |     |             |
|------------------|--|-------|------|-----|-------------|
| RESULT 1         |  |       |      |     |             |
| LOCUS BE386120/c |  |       |      |     |             |
| DEFINITION       | 601277233F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3617961 5', mRNA sequence.  | 40 bp | mRNA | EST | 21-JUL-2000 |
| ACCESSION        | BE386120   |       |      |     |             |
| VERSION          | BE386120.1   |       |      |     |             |
| KEYWORDS         | EST.   |       |      |     |             |
| SOURCE           | human.   |       |      |     |             |
| ORGANISM         | Homo sapiens   |       |      |     |             |
| REFERENCE        | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |       |      |     |             |
| AUTHORS          | 1 (bases 1 to 40)  |       |      |     |             |
| JOURNAL          | NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .  |       |      |     |             |
| COMMENT          | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D.          |       |      |     |             |

Email: c9abps-r@mail.nih.gov  
Tissue Procurement: ATCC/DCMD/BTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: L1CM286 row: b column: 10.  
Location/Qualifiers  
1. .40

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3617961"  
/clone\_lib="NIH\_MGC\_20"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 12 a 8 c 11 g 9 t  
ORIGIN

Query Match 100.0%; Score 9; DB 167; Length 40;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taataactag 9  
|||||  
Db 19 TAATACTAG 11

RESULT 2

AZ769245 55 bp DNA GSS 16-FEB-2001  
LOCUS IM0569010R Mouse 10kb plasmid UUGCIM library Mus musculus genomic clone UUGCIM0569010 R, DNA sequence.  
ACCESSION AZ769245  
VERSION AZ769245.1 GI:12889180  
KEYWORDS GSS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 55)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D.,Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0369 row: 0 column: 10  
Seq primer: CACACAGGAACACGATGACC  
Class: plasmid ends  
High quality sequence stop: 55.  
Location/Qualifiers  
1. .55  
/organism="Mus musculus"

FEATURES  
source

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0569010"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      33 a      4 c      4 g      14 t
ORIGIN

Query Match      100.0%; Score 9; DB 249; Length 55;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 10 TAATACTAG 18

RESULT 3
AF097417 57 bp mRNA EST 10-NOV-1999
LOCUS AF097417 Vespertilio superans library Vespertilio superans cDNA, mRNA sequence.
DEFINITION AF097417
ACCESSION AF097417.1 GI:6341356
VERSION AF097417.1
KEYWORDS EST.
SOURCE Asian particolored bat.
ORGANISM Vespertilio superans
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Chiroptera; Microchiroptera; Vespertilionidae; Vesperillo.
REFERENCE 1 (bases 1 to 57)
AUTHORS Kang,H.S. and Kim,J.W.
TITLE Differential Gene Expression of Active and Hibernating States in Bats
JOURNAL Unpublished (1999)
COMMENT Contact: Kang HS
Anatomy
Wonju College of Medicine, Yonsei University
Ilsandong 162, Wonju, Kangwon 220-701, Korea
Email: mhkim@yunc.yonsei.ac.kr.
FEATURES
Location/Qualifiers
1..57
/organism="Vesperillo superans"
/db_xref="taxon:105273"
/clone_lib="Vesperillo superans library"
/note="country=South Korea; Kangwon Province"
BASE COUNT      19 a      9 c      9 g      20 t
ORIGIN

Query Match      100.0%; Score 9; DB 14; Length 57;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 42 TAATACTAG 50

RESULT 4
AF097417 57 bp DNA GSS 03-OCT-2000
LOCUS AF097417 Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0220E09 R, DNA sequence.
DEFINITION AF097417
ACCESSION AF097417.1 GI:10558315
VERSION AF097417.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 57)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0220 row: E column: 09
Seq primer: CACACAGGAACACGATGACC
Class: plasmid ends
High quality sequence stop: 57.
Location/Qualifiers
1..57
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0220E09"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      12 a      14 c      9 g      22 t
ORIGIN

```

```

Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taactag 9
DB 40 TAATACTAG 32

RESULT 5
AA637860/c
LOCUS
DEFINITION vt29f01.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
IMAGE:1122073 5' similar to SW:CYB_MOUSE P00158 CYTOCHROME B ;
mRNA sequence.
ACCESSION AA637860
VERSION AA637860
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286-1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:611409
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..58
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:1122073"
/clone_lib="Barstead mouse myotubes MPLRB5"
/cell_line="C2C12"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(GT) primer [5',
TCTTACGAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified p7T3 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."
BASE COUNT 17 a 14 c 10 g 17 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 58;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taactag 9
DB 13 TAATACTAG 5

RESULT 7
AA215820/c
LOCUS
DEFINITION AW215820
mRNA sequence.
ACCESSION AW215820
VERSION AW215820.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 59)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Db 28 TAATACTAG 20

RESULT 6
AW215812/c
LOCUS
DEFINITION AW215812
mRNA sequence.
ACCESSION AW215812
VERSION AW215812.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 59)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other_ESTs: uo99e06.y1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
MGI:1031142
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..59
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:2650690"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; transgenic model WNT-1, expression driven by
MTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 14 a 11 c 7 g 27 t
ORIGIN

Query Match 100.0%; Score 9; DB 113; Length 59;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taactag 9
DB 13 TAATACTAG 5

RESULT 7
AA215820/c
LOCUS
DEFINITION AW215820
mRNA sequence.
ACCESSION AW215820
VERSION AW215820.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 59)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

AUTHORS  
TITLE  
JOURNAL  
COMMENT

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps@email.nih.gov](mailto:cgaps@email.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

FEATURES  
source

1. .59  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2650691"  
/clone\_lib="NCI\_CGAP\_Lu30"  
/tissue\_type="tumor, metastatic to mammary"  
/lab\_host="DHL08"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; transgenic model WNT-1, expression driven by  
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo  
dT. Library constructed by Life Technologies.  
Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT  
ORIGIN

14 a 11 c 7 g 27 t

Query Match 100.0%; Score 9; DB 113; Length 59;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9  
|||||||

Db 13 TAATACTAG 5

RESULT 8  
TAB2F10Q  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

TAB2F10Q 62 bp DNA GSS 13-DEC-2000  
T. brucei sheared genomic DNA clone 82f10, reverse sequence,  
genomic survey sequence.  
AL459971 GI:11860296  
GSS.  
Trypanosoma brucei.  
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 62)  
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,  
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,  
Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: [barrell@sanger.ac.uk](mailto:barrell@sanger.ac.uk) and  
[nhs@sanger.ac.uk](mailto:nhs@sanger.ac.uk)  
Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
to give a tight size distribution (8-10 kb). The v + i method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).

FEATURES  
source

1. .62  
Location/Qualifiers  
/organism="Trypanosoma brucei"  
/strain="TREU927"  
/db\_xref="taxon:5691"  
/clone="82f10"

BASE COUNT  
ORIGIN

27 a 8 c 11 g 16 t

Query Match 100.0%; Score 9; DB 258; Length 62;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9  
|||||||

Db 10 TAATACTAG 18

RESULT 9  
AZ808515/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AZ808515 66 bp DNA GSS 20-FEB-2001  
2M0072A07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0072A07 F, DNA sequence.  
AZ808515  
GSS.  
AZ808515.1 GI:12973933  
GSS.  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 66)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0072 row: A column: 07  
Seq primer: CGTTGTAACGACGACGACGAGT  
Class: plasmid ends  
High quality sequence stop: 66.  
Location/Qualifiers  
1. .66  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0072A07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The

FEATURES  
source

1. .66  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0072A07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The



Thu Oct 25 13:08:55 2001

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:1081639  
 Seq primer: -40RP from Gibco.

FEATURES  
 source  
 1. .71  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3372035"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 32 a 10 c 12 g 17 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9  
 |||||  
 Db 58 TAATACTAG 66

RESULT 14  
 BE630111 71 bp mRNA EST 25-AUG-2000  
 LOCUS u15f06.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 DEFINITION IMAGE:3372035 5', mRNA sequence.  
 BE630111  
 BE630111.1 GI:9912799  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:1081639.

FEATURES  
 source  
 1. .71  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3372035"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 32 a 10 c 12 g 17 t  
 ORIGIN

Query Match 100.0%; Score 9; DB 123; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9  
 |||||  
 Db 58 TAATACTAG 66

RESULT 13  
 BE627067 71 bp mRNA EST 24-AUG-2000  
 LOCUS u15f06.y2 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 DEFINITION IMAGE:3372035 5', mRNA sequence.  
 BE627067  
 BE627067.1 GI:9907487  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.



gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I, and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 17 a 12 c 11 g 31 t  
ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
|||||  
Db 13 TAATACTAG 5

RESULT 15  
BE688187/c  
LOCUS  
DEFINITION BE688187 71 bp mRNA EST 11-SEP-2000  
IMAGE:3464723 3', mRNA sequence.  
ACCESSION BE688187  
VERSION  
KEYWORDS  
SOURCE EST.  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 71)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MG1:1384083.

FEATURES  
Source  
1..71  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3464723"  
/clone\_lib="Soares.mammary.gland\_NMLMG".  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I, and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 18 a 12 c 9 g 32 t  
ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
|||||  
Db 13 TAATACTAG 5

us-09-462-955-4.rst

Thu Oct 25 13:08:55 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:31 ; Search time 180.6 Seconds  
(without alignments)  
9.434 Million cell updates/sec

Title: US-09-462-955-4

Perfect score: 9  
Sequence: 1 taactag 9

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description       |
|------------|-------|---------------|--------|-------|-------------------|
| C 1        | 9     | 100.0         | 12     | 2     | US-08-418-071-15  |
| C 2        | 9     | 100.0         | 30     | 3     | US-09-150-133-33  |
| C 3        | 9     | 100.0         | 30     | 3     | US-09-150-141-33  |
| C 4        | 9     | 100.0         | 30     | 4     | US-09-374-493-33  |
| C 5        | 9     | 100.0         | 30     | 4     | US-09-374-824-33  |
| C 6        | 9     | 100.0         | 30     | 4     | US-09-374-492-33  |
| C 7        | 9     | 100.0         | 38     | 3     | US-08-817-926-41  |
| C 8        | 9     | 100.0         | 41     | 3     | US-08-491-954-106 |
| C 9        | 9     | 100.0         | 62     | 3     | US-08-492-076-9   |
| C 10       | 9     | 100.0         | 78     | 2     | US-08-680-326-51  |
| C 11       | 9     | 100.0         | 101    | 1     | US-07-795-859B-14 |
| C 12       | 9     | 100.0         | 101    | 1     | US-08-616-14      |
| C 13       | 9     | 100.0         | 299    | 2     | US-08-637-759B-31 |
| C 14       | 9     | 100.0         | 299    | 3     | US-08-871-355A-31 |
| C 15       | 9     | 100.0         | 352    | 6     | 5242821-11        |
| C 16       | 9     | 100.0         | 454    | 1     | US-08-846-134-2   |
| C 17       | 9     | 100.0         | 492    | 1     | US-08-470-720-3   |
| C 18       | 9     | 100.0         | 573    | 1     | US-08-709-912-18  |
| C 19       | 9     | 100.0         | 573    | 2     | US-09-047-370-18  |
| C 20       | 9     | 100.0         | 599    | 3     | US-08-705-875A-2  |
| C 21       | 9     | 100.0         | 600    | 5     | PCT-US91-02766-24 |
| C 22       | 9     | 100.0         | 628    | 2     | US-08-874-186-45  |
| C 23       | 9     | 100.0         | 660    | 3     | US-09-284-782-15  |
| C 24       | 9     | 100.0         | 663    | 4     | US-08-998-416-187 |
| C 25       | 9     | 100.0         | 706    | 3     | US-08-654-025-4   |
| C 26       | 9     | 100.0         | 713    | 1     | US-08-234-939-9   |
| C 27       | 9     | 100.0         | 713    | 1     | US-08-558-865-9   |

Sequence 849, Appl  
Sequence 1138, Appl  
Sequence 628, Appl  
Sequence 261, Appl  
Sequence 11, Appl  
Sequence 56, Appl  
Sequence 50, Appl  
Sequence 289, Appl  
Sequence 8, Appl  
Sequence 3, Appl  
Sequence 47, Appl  
Sequence 9, Appl  
Sequence 48, Appl  
Sequence 1, Appl

US-08-418-071-15/c  
; Sequence 15, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh  
; APPLICANT: You, Li-Ru  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/418,071  
; FILING DATE: 06-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: DCB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; MOLECULE TYPE: other nucleic acid  
; ANTI-SENSE: NO  
US-08-418-071-15

## ALIGNMENTS

RESULT 1  
US-08-418-071-15/c  
; Sequence 15, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh  
; APPLICANT: You, Li-Ru  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/418,071  
; FILING DATE: 06-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: DCB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; MOLECULE TYPE: other nucleic acid  
; ANTI-SENSE: NO  
US-08-418-071-15

Query Match 100.0%; Score 9; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu Oct 25 13:08:53 2001

```
Qy 1 taatactag 9
    |||||
Db 10 TAATACTAG 2

RESULT 2
US-09-150-133-33/c
; Sequence 33, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-150-133-33

Query Match 100.0%; Score 9; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 3
US-09-150-141-33/c
; Sequence 33, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-150-141-33

Query Match 100.0%; Score 9; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 4
US-09-374-493-33/c
; Sequence 33, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
```

```
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-493-33

Query Match 100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 5
US-09-374-824-33/c
; Sequence 33, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-824-33

Query Match 100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 6
US-09-374-492-33/c
; Sequence 33, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
```

```

; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-492-33

```

```

Query Match      100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 taatactag 9
Db 12 TAATACTAG 4

```

```

RESULT 7
US-08-817-926-41
; Sequence 41, Application US/08817926
; Patent No. 6001590
; GENERAL INFORMATION:
; APPLICANT: Kameda, Toshihiro
; APPLICANT: Suda, Hisako
; APPLICANT: Tamai, Yukio
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Kato, No. 6001590no
; APPLICANT: Sakai, Yasuyoshi
; TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINII
; TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,926
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/02597
; FILING DATE: 12-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 234133/1995
; FILING DATE: 12-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 42536/1996
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 081356/0112
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
US-08-817-926-41

```

```

Query Match      100.0%; Score 9; DB 3; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 taatactag 9
Db 10 TAATACTAG 18

```

```

RESULT 8
US-08-491-954-106/c
; Sequence 106, Application US/08491954
; Patent No. 6096321
; GENERAL INFORMATION:
; APPLICANT: Girardeau, Jean-Pierre
; APPLICANT: Martin, Christine
; APPLICANT: Mechlin, Marie-Claire
; APPLICANT: Der Vartanian, Maurice
; APPLICANT: Bousquet, Francois
; TITLE OF INVENTION: SUB-UNIT OF CS31A PROTEIN CAPSULE
; TITLE OF INVENTION: MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A
; TITLE OF INVENTION: PROTEIN CAPSULE INCLUDING SUCH A SUB-UNIT, AND
; TITLE OF INVENTION: MICROORGANISMS WHOSE OUTER MEMBRANE CARRIES SUCH
; TITLE OF INVENTION: SUB-UNITS, AND PROCEDURE FOR OBTAINING AND UTILIZING SUCH
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,954
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/01281
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989,6264P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna

```

```
;
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..41
;
; US-08-491-954-106
;
;
; Query Match      100.0%; Score 9; DB 3; Length 41;
; Best Local Similarity 100.0%; Pred. No. 1.4e+03;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 taatactag 9
;      |||||
; Db 9 TAATACTAG 1
;
;
; RESULT 9
; US-08-492-076-9/c
; ; Sequence 9, Application US/08492076A
; ; Patent No. 6060064
; ; GENERAL INFORMATION:
; ; APPLICANT: Adams, Sally E.
; ; APPLICANT: Burus, Nigel R.
; ; APPLICANT: Richardson, Simon M.
; ; TITLE OF INVENTION: No. 6060064el Proteinaceous Particles
; ; FILE REFERENCE: 10180.60968
; ; CURRENT APPLICATION NUMBER: US/08/492,076A
; ; CURRENT FILING DATE: 1995-06-28
; ; EARLIER APPLICATION NUMBER: PCT/GB93/02656
; ; EARLIER FILING DATE: 1993-12-24
; ; NUMBER OF SEQ ID NOS: 23
; ; SOFTWARE: Patentin Ver. 2.0
; ; SEQ ID NO 9
; ; LENGTH: 62
; ; TYPE: DNA
; ; ORGANISM: Artificial Sequence
; ; FEATURE:
; ; OTHER INFORMATION: Description of Artificial Sequence: Partial V3
; ; OTHER INFORMATION: loop of HIV-1 further comprising restriction
; ; OTHER INFORMATION: enzyme sites.
;
; US-08-492-076-9
;
; Query Match      100.0%; Score 9; DB 3; Length 62;
; Best Local Similarity 100.0%; Pred. No. 1.3e+03;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 taatactag 9
;      |||||
; Db 9 TAATACTAG 1
;
;
; RESULT 10
; US-08-680-326-51/c
; ; Sequence 51, Application US/08680326
; ; Patent No. 5925733
; ; GENERAL INFORMATION:
; ; APPLICANT: ROSE, TIMOTHY M.
; ; APPLICANT: BOSCH, MARIX
; ; APPLICANT: STRAND, KURT
; ; APPLICANT: TODARO, GEORGE J.
; ; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
; ; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
; ; TITLE OF INVENTION: FIBROMATOSIS
; ; NUMBER OF SEQUENCES: 152
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: MORRISON & FORSTER
; ; STREET: 755 Page Mill Road
; ; CITY: Palo Alto
; ; STATE: California
; ; COUNTRY: USA
; ; ZIP: 94304-1018
; ; COMPUTER READABLE FORM: disk
; ; MEDIUM TYPE: Floppy
```

```
;
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680.326
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Schlif, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20001.00
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-680-326-51
;
; Query Match      100.0%; Score 9; DB 2; Length 78;
; Best Local Similarity 100.0%; Pred. No. 1.2e+03;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 taatactag 9
;      |||||
; Db 37 TAATACTAG 29
;
;
; RESULT 11
; US-07-795-859B-14
; ; Sequence 14, Application US/07795859B
; ; Patent No. 5422262
; ; GENERAL INFORMATION:
; ; APPLICANT: Anderson, Stefan
; ; APPLICANT: Russell, David W.
; ; TITLE OF INVENTION: Steroid 5'-Reductases
; ; NUMBER OF SEQUENCES: 37
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Arnold, White and Durkee
; ; STREET: P.O. Box 4433
; ; CITY: Houston
; ; STATE: TX
; ; COUNTRY: USA
; ; ZIP: 77210
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: ASCII-DOS
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/07/795,859B
; ; FILING DATE: 18-NOV-1991
; ; CLASSIFICATION: 435
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Parker, David L.
; ; REGISTRATION NUMBER: 32,165
; ; REFERENCE/DOCKET NUMBER: UTSD:260/PAR
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (512) 320-7200
; ; TELEFAX: (512) 474-7677
; ; INFORMATION FOR SEQ ID NO: 14:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 101 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: DNA (genomic)
;
; US-07-795-859B-14
```

Query Match 100.0%; Score 9; DB 1; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 |||||  
 Db 18 TAATACTAG 26

## RESULT 12

US-08-457-616-14  
 ; Sequence 14, Application US/08457616  
 ; Patent No. 5679521  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Anderson, Stefan  
 ; APPLICANT: Russell, David W.  
 ; TITLE OF INVENTION: Steroid 5 $\alpha$ -Reductases  
 ; NUMBER OF SEQUENCES: 37  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold, White and Durkee  
 ; STREET: P O Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII-DOS  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/457,616  
 ; FILING DATE: 01-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/795,859  
 ; FILING DATE: 18-NOV-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Parker, David L.  
 ; REGISTRATION NUMBER: 32,165  
 ; REFERENCE/DOCKET NUMBER: UTSD:260/PAR  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512) 320-7200  
 ; TELEFAX: (512) 474-7677  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 101 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-457-616-14

Query Match 100.0%; Score 9; DB 1; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 |||||  
 Db 18 TAATACTAG 26

## RESULT 13

US-08-637-759B-31  
 ; Sequence 31, Application US/08637759B  
 ; Patent No. 5876931  
 ; GENERAL INFORMATION:  
 ; APPLICANT: David William Holden  
 ; TITLE OF INVENTION: Identification of Genes  
 ; NUMBER OF SEQUENCES: 501

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Patrea L. Pabst  
 ; STREET: 2800 One Atlantic Center  
 ; STREET: 1201 West Peachtree Street  
 ; CITY: Atlanta  
 ; STATE: Georgia  
 ; COUNTRY: USA  
 ; ZIP: 30309-3450  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/637,759B  
 ; FILING DATE: 03-MAY-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/GB95/02875  
 ; FILING DATE: 11-DEC-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Pabst, Patrea L.  
 ; REGISTRATION NUMBER: 31,284  
 ; REFERENCE/DOCKET NUMBER: RPMS 101  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (404) 873-8794  
 ; TELEFAX: (404) 873-8795  
 ; INFORMATION FOR SEQ ID NO: 31:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 299 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Partial sequence of Salmonella typhimurium  
 ; ORGANISM: virulence gene  
 ; US-08-637-759B-31

Query Match 100.0%; Score 9; DB 2; Length 299;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9  
 |||||  
 Db 44 TAATACTAG 52

## RESULT 14

US-08-871-355A-31  
 ; Sequence 31, Application US/08871355A  
 ; Patent No. 6015669  
 ; GENERAL INFORMATION:  
 ; APPLICANT: David William Holden  
 ; TITLE OF INVENTION: Identification of Genes  
 ; NUMBER OF SEQUENCES: 501  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Patrea L. Pabst  
 ; STREET: 2800 One Atlantic Center  
 ; STREET: 1201 West Peachtree Street  
 ; CITY: Atlanta  
 ; STATE: Georgia  
 ; COUNTRY: USA  
 ; ZIP: 30309-3450  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

Thu Oct 25 13:08:53 2001

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871.355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
US-08-871-355A-31

```

```

Query Match      100.0%; Score 9; DB 3; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.le+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 taatactag 9
    |||||
Db 44 TAATACTAG 52

```

```

RESULT 15
5242821-11/c
; Patent No. 5242821
; APPLICANT: PALVA, LIKKA; SIBAKOV, MERV
; TITLE OF INVENTION: LACTOCOCCUS PROMOTER AND SIGNAL
; SEQUENCES FOR EXPRESSION IN BACTERIA
; NUMBER OF SEQUENCES: 27
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/377,450
; FILING DATE: 10-JUL-1989
; SEQ ID NO: 11
; LENGTH: 352
5242821-11

```

```

Query Match      100.0%; Score 9; DB 6; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.le+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 taatactag 9
    |||||
Db 139 TAATACTAG 131

```

Search completed: October 24, 2001, 10:00:31  
Job time: 216 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:28 ; Search time 393.52 seconds  
(without alignments)  
14.360 Million cell updates/sec

Title: US-09-462-955-5

Perfect score: 9

Sequence: 1 ctatgatta 9

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL1/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description  |
|------------|-------|-------------|--------|----------|--|
| 1          | 9     | 100.0       | 18     | AAV33149 | Plasmodium falciparum                                    |
| 2          | 9     | 100.0       | 18     | AAV33146 | Plasmodium vivax   |
| 3          | 9     | 100.0       | 22     | AA30640  | Human G protein-coupled receptor 1 to 3' PCR primer-1 to |
| 4          | 9     | 100.0       | 22     | AA250574 | Double-stranded DN                                       |
| 5          | 9     | 100.0       | 22     | AA31603  | Human NKCC2 gene   |
| 6          | 9     | 100.0       | 23     | AAV45715 | Reverse PCR primer                                       |
| 7          | 9     | 100.0       | 28     | AA261141 | Human TPST-1 cDNA  |
| 8          | 9     | 100.0       | 30     | AA261141 | Human TPST-1 cDNA  |
| 9          | 9     | 100.0       | 30     | AA261141 | Human TPST-1 cDNA  |
| 10         | 9     | 100.0       | 30     | AA261141 | Human TPST-1 cDNA  |
| 11         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 12         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 13         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 14         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 15         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 16         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 17         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 18         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 19         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 20         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 21         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |
| 22         | 9     | 100.0       | 31     | AA261141 | Human TPST-1 cDNA  |

|   |    |   |       |     |    |           |                      |
|---|----|---|-------|-----|----|-----------|----------------------|
| C | 12 | 9 | 100.0 | 31  | 20 | AA02686   | Gemini virus DNA f   |
|   | 13 | 9 | 100.0 | 32  | 18 | AA090052  | Primer for chicken   |
|   | 14 | 9 | 100.0 | 35  | 18 | AAT77100  | Pisum sativum plas   |
|   | 15 | 9 | 100.0 | 35  | 21 | AA039333  | Z. paucivorans 16S   |
|   | 16 | 9 | 100.0 | 36  | 20 | AA028348  | PCR primer PC641 f   |
| C | 17 | 9 | 100.0 | 38  | 18 | AAT62191  | PCR primer PF668 f   |
|   | 18 | 9 | 100.0 | 38  | 20 | AAT289335 | PCR primer PF668 f   |
| C | 19 | 9 | 100.0 | 38  | 21 | AAZ61496  | Primer PCR2AS for    |
|   | 20 | 9 | 100.0 | 38  | 22 | AAF91511  | Primer PCR2AS for    |
|   | 21 | 9 | 100.0 | 38  | 22 | AAF91519  | Primer PCR2AS for    |
| C | 22 | 9 | 100.0 | 39  | 21 | AAA30642  | Primer PCR2AS for    |
| C | 23 | 9 | 100.0 | 39  | 21 | AAZ50576  | Human G protein-co   |
|   | 24 | 9 | 100.0 | 39  | 22 | AAF91503  | 3' PCR primer-2 to   |
|   | 25 | 9 | 100.0 | 41  | 15 | AAQ79998  | Human gene/PoRa prom |
| C | 26 | 9 | 100.0 | 53  | 21 | AAA48234  | Linker coding for    |
|   | 27 | 9 | 100.0 | 62  | 15 | AAQ69204  | T. reesei xylanase   |
|   | 28 | 9 | 100.0 | 67  | 21 | AAA48233  | Oligo encoding N-t   |
|   | 29 | 9 | 100.0 | 74  | 22 | AAF31601  | T. reesei xylanase   |
|   | 30 | 9 | 100.0 | 93  | 16 | AAT26341  | Variable oligonucl   |
|   | 31 | 9 | 100.0 | 98  | 21 | AAC09195  | Human gene signatu   |
| C | 32 | 9 | 100.0 | 101 | 16 | AAQ97387  | Human secreted pro   |
| C | 33 | 9 | 100.0 | 110 | 18 | AAV75998  | Human type I stereo  |
| C | 34 | 9 | 100.0 | 112 | 21 | AAZ17499  | Staphylococcus aur   |
| C | 35 | 9 | 100.0 | 114 | 21 | AAZ94806  | Human secreted pro   |
|   | 36 | 9 | 100.0 | 116 | 21 | AAC10236  | Soybean microsate    |
| C | 37 | 9 | 100.0 | 127 | 18 | AAV76841  | Human secreted pro   |
|   | 38 | 9 | 100.0 | 138 | 21 | AAC18454  | Human secreted pro   |
|   | 39 | 9 | 100.0 | 138 | 21 | AAC20546  | Human secreted pro   |
| C | 40 | 9 | 100.0 | 149 | 21 | AAA45738  | Human secreted pro   |
| C | 41 | 9 | 100.0 | 151 | 21 | AAC26107  | Human secreted exp   |
| C | 42 | 9 | 100.0 | 152 | 16 | AAT23458  | Human secreted pro   |
|   | 43 | 9 | 100.0 | 159 | 21 | AA445500  | Human gene signatu   |
|   | 44 | 9 | 100.0 | 165 | 21 | AAC22885  | Human secreted exp   |
| C | 45 | 9 | 100.0 | 177 | 19 | AA010779  | Human secreted pro   |
|   |    |   |       |     |    |           | Human biallelic pro  |

#### ALIGNMENTS

|          |   |
|----------|---|
| RESULT   | 1   |
| AAV33149 | AAV33149 standard; DNA; 18 BP.                                    |
| XX       | AAV33149  |
| AC       | AAV33149;   |
| XX       | 07-DEC-1998 (first entry)   |
| DT       | Plasmodium falciparum coxI gene PCR primer PFL.                   |
| DE       | Malaria; infection; therapy; diagnosis; mitochondrion; coxI gene; |
| XX       | cytochrome C oxidase; PCR; primer; ss.                            |
| KW       | Synthetic.  |
| XX       | Plasmodium falciparum.  |
| OS       | WO9835057-A1.   |
| PN       | 13-AUG-1998.  |
| XX       | 05-FEB-1998; 98WO-IB00212.  |
| PD       | 26-SEP-1997; 97AU-0009481.  |
| PF       | 06-FEB-1997; 97AU-0004953.  |
| XX       | 21-APR-1997; 97AU-0006329.  |
| XX       | (MOLE-) INST MOLECULAR & CELL BIOLOGY.                            |
| PA       | (OYSL-) UNIV SINGAPORE NAT.                                       |
| XX       | Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;                  |
| PI       | WPI; 1998-447251/38.  |
| DR       | Detecting Plasmodium infection from hybridisation with            |
| XX       |   |
| XX       |   |
| PT       |   |

PT extrachromosomal element - providing genus or species specific  
 PT diagnosis with few false negatives, in humans or animals

XX Claim 6; Page 66; 120pp; English.

XX This synthetic oligonucleotide sequence (PFI) is derived from the  
 CC Plasmodium vivax mitochondrial cytochrome C oxidase (coxI) gene.  
 CC In a claimed method, Plasmodium is detected in a human or animal  
 CC sample (especially blood or dried blood) by treating it, or  
 CC animal sample (especially blood or dried blood) by treating it, or  
 CC derived nucleic acid, with a Plasmodium extrachromosomal genetic  
 CC element or derived nucleic acid (A) and detecting any hybridisation.  
 CC (A) can include the PSI-PLA470, PLH-PPH, PRB or PWQ gene (see  
 CC AAV33135-38), the coxI gene or nucleic acids derived from them. Also  
 CC new are (A)-specific probes and primers (see AAV33139-56). The method  
 CC is used to diagnose Plasmodium infection. The high degree of  
 CC similarity between (A) from different species allows development of  
 CC genus- or species-specific assays that result in fewer false  
 CC negatives than known methods (typically 1% against 3%). Primer PFI  
 CC can be used with another primer (see AAV33145) for the species-  
 CC specific detection of P. falciparum.

XX Sequence 18 BP; 4 A; 3 C; 3 G; 8 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;

Best Local Similarity 100.0%; Pred. No. 3.8e+03; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatgattata 9

Db 2 ctatgattata 10

RESULT 2

AAV33146

ID AAV33146 standard; DNA; 18 BP.

XX AAV33146;

XX 07-DEC-1998 (first entry)

Plasmodium vivax coxI gene PCR primer PFI.

Malaria; infection; therapy; diagnosis; mitochondrion; coxI gene;

cytochrome C oxidase; PCR; primer; ss.

Synthetic.

Plasmodium vivax.

WO9835057-A1.

13-AUG-1998.

05-FEB-1998; 98WO-IB00212.

26-SEP-1997; 97AU-0009481.

06-FEB-1997; 97AU-0004953.

21-APR-1997; 97AU-0006329.

(MOLE-) INST MOLECULAR & CELL BIOLOGY.

(UYSI-) UNIV SINGAPORE NAT.

Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;

WPI; 1998-447251/38.

Detecting Plasmodium infection from hybridisation with

extrachromosomal element - providing genus or species specific

diagnosis with few false negatives, in humans or animals

Claim 6; Page 65; 120pp; English.

This synthetic oligonucleotide sequence (PFI) is derived from the

CC Plasmodium vivax mitochondrial cytochrome C oxidase (coxI) gene.  
 CC In a claimed method, Plasmodium is detected in a human or animal  
 CC sample by treating it, or derived nucleic acid, with a Plasmodium  
 CC extrachromosomal genetic element or derived nucleic acid (A) and  
 CC detecting any hybridisation. (A) can include the PSI-PLA470,  
 CC PLH-PPH, PRB or PWQ gene (see AAV33135-38), the coxI gene or nucleic  
 CC acids derived from them. Also new are (A)-specific probes and  
 CC primers (see AAV33139-56). The method is used to diagnose Plasmodium  
 CC infection. The sample can be blood or dried blood. The high  
 CC degree of similarity between (A) from different species allows  
 CC development of genus- or species-specific assays that result in  
 CC fewer false negatives than known methods (typically 1% against 3%).  
 CC Primer PFI can be used with another primer (see AAV33145) for the  
 CC species-specific detection of P. vivax.

XX Sequence 18 BP; 3 A; 3 C; 3 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;

Best Local Similarity 100.0%; Pred. No. 3.8e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatgattata 9

Db 2 ctatgattata 10

RESULT 3

AAA30640/C

ID AAA30640 standard; DNA; 22 BP.

XX AAA30640;

XX 21-AUG-2000 (first entry)

Human G protein-coupled receptor GHSR PCR primer, SEQ ID NO:84.

G protein-coupled receptor; GPCR; constitutively active;

intracellular loop 3; transmembrane domain 6; drug screening;

agonist; antagonist; PCR primer; ss.

Homo sapiens.

WO200022129-A1.

20-APR-2000.

12-OCT-1999; 99WO-US23938.

13-OCT-1998; 98US-0170496.

(AREN-) ARENA PHARM INC.

Behan DP, Chalmers DT, Liaw CW;

WPI; 2000-329165/28.

Non-endogenous constitutively activated human G protein-coupled

receptors, useful for identifying agonists for use as pharmaceutical

agents

Example 1; Page 42; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions  
 CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAV90643-  
 CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAV30709-A30743  
 CC and AAA30775-A30779). The mutant proteins of the invention contain a  
 CC mutation in a portion of the protein comprising intracellular loop 3  
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,  
 CC is substituted for an endogenous residue in IC3 at a position 16 amino  
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence  
 CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg  
 CC or Ala, and is preferably Lys. When the endogenous residue at this

CC position is Lys, this residue is replaced by His, Arg or preferably Ala.  
 CC The 15 amino acid stretch between the substituted amino acid and the Pro  
 CC may be endogenous, non-endogenous, or a mixture of endogenous and  
 CC non-endogenous residues. The constitutively active GPCRs are useful for  
 CC identifying antagonists, agonists and partial agonists for use as  
 CC pharmaceutical agents. The mutant proteins are also useful in research  
 CC settings for elucidating the roles of the receptors in normal and  
 CC diseased conditions. Antagonists for a particular GPCR are useful for  
 CC treating diseases and disorders associated with that receptor. Because  
 CC the novel mutant GPCRs are constitutively active, they can be used  
 CC directly for screening of compounds without the need for endogenous  
 CC ligands. The present sequence represents a PCR primer used in an  
 CC exemplification of the invention to isolate cDNA encoding a human  
 CC wild-type GPCR for cloning, an exemplification of the invention. The GPCR  
 CC cDNA was subjected to site-directed mutagenesis (SDM) to generate DNA  
 CC encoding the corresponding mutant of the invention.

CC Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9  
 Db 17 CTAGTATTATA 9

RESULT 4  
 AA250574/C  
 ID AA250574 standard; DNA; 22 BP.

XX AA250574;

XX 20-JUN-2000 (first entry)

XX 3' PCR primer-1 to amplify human GHSR DNA.

XX G protein-coupled orphan receptor; GPCR; agonist; G protein;  
 KW GPCR fusion protein; inverse agonist; drug; treatment; PCR primer;  
 KW GHSR; G protein-coupled receptor; human; ss.

XX Homo sapiens.

XX WO200006597-A2.

XX 10-FEB-2000.

XX 30-JUL-1999; 99WO-US17425.

XX 31-JUL-1998; 98US-0094879.

XX 30-OCT-1998; 98US-0106300.

XX 04-DEC-1998; 98US-0110906.

XX 26-FEB-1999; 99US-0121851.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw C, Lin I, Lowitz K, Chen R;

XX WPI; 2000-195260/17.

XX Identification of a compound useful as a therapeutic agent, comprises

XX identifying a compound against constitutively activated G

XX protein-coupled orphan receptors -

XX Example 2; Page 27; 123pp; English.

XX The patent discloses a method of identifying agonists and inverse or  
 CC partial agonists to the endogenous, constitutively activated  
 CC G protein-coupled orphan receptors (GPCRs), by contacting them with a  
 CC GPCR fusion protein comprising a GPCR and a G protein. Determining  
 CC expression of GPCRs in tissue samples can be used to identify related

CC diseases. Inverse agonists to these receptors can be used as drugs for  
 CC treating GPCR-related diseases. The present sequence is a 3' PCR primer,  
 CC used in primary PCR reaction to amplify human GHSR DNA from  
 CC hippocampus cDNA template.

CC Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9  
 Db 17 CTAGTATTATA 9

RESULT 5  
 AAF31603/C

ID AAF31603 standard; DNA; 22 BP.

XX AAF31603;

XX 09-APR-2001 (first entry)

XX Double-stranded DNA tag PCR primer pTag2.

XX attenuated microorganism; signature tagged transposon mutant;  
 KW mutant library; mycobacterial infection; actinomycetales;  
 KW antibacterial; immunostimulant; vaccine; PCR primer; ss.

XX Unidentified.

XX WO200102555-A1.

XX 11-JAN-2001.

XX 06-JUL-2000; 2000WO-IB00950.

XX 06-JUL-1999; 99US-0142982.

XX 08-JUL-1999; 99US-0142833.

XX (INSP ) INST PASTEUR.

XX Gicquel B, Guilhot C, Camacho L;

XX WPI; 2001-091804/10.

XX Screening a mutant library for mutants unable to grow under specific  
 PT conditions and for identifying loci involved in pathogenicity,  
 PT comprises using signature tagged transposon mutagenesis -

XX Example 4; Page 19; 159pp; English.

XX The present sequence is given in a specification relating to a method for  
 CC screening a library of mutants. The method comprises constructing a  
 CC library with insertions in genes and/or regulatory regions of the  
 CC organisms of interest, where the insertion contains a tag and/or a  
 CC transposon associated with a tag. The mutants are identified by  
 CC hybridisation of the tags to known sequences. The method is useful for  
 CC treating an individual suffering from a mycobacterial infection,  
 CC suspected of being infected with a Mycobacterium, or having been  
 CC exposed to an infectious Mycobacterium. It is also useful for  
 CC identifying and isolating mutants of Mycobacterium. It is also useful for  
 CC compounds that have antibiotic activity. The method is used to identify  
 CC mutants of microorganisms, preferably an actinomycetales, such as  
 CC M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and  
 CC M. paratuberculosis, that is unable to grow under specific conditions.  
 CC It is especially useful for identifying loci involved in pathogenicity.  
 CC It is useful in constructing vaccines. The method can be used to screen  
 CC multiple libraries concurrently. It can screen libraries of different  
 CC organisms or different strains of the same organism. The present  
 CC sequence was used to create the tag used in the construction of tagged

CC mutants.  
XX Sequence 22 BP; 7 A; 5 C; 1 G; 9 T; 0 other;  
SQ

Query Match 100.0%; Score 9; DB 22; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9  
DB 11 CTAGTATTA 3

RESULT 6  
AAV45715/C  
ID AAV45715 standard; DNA; 23 BP.  
XX  
AC AAV45715;  
XX  
DT 21-DEC-1998 (first entry)  
XX  
DE Human NKCC2 gene exon 19 forward primer hNKCC2ex19.  
XX  
XX Na-K-2Cl cotransporter; NKCC2; human; Bartter's syndrome;  
KW ion transport; hypokalaemic alkalosis; hypercalciuria;  
KW nephrocalcinosis; diagnosis; therapy; SSCP; primer; ss.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO9829431-A1.  
XX  
PD 09-JUL-1998.  
XX  
XX 19-DEC-1997; 97WO-US23553.  
PF  
XX 31-DEC-1996; 96US-0778052.  
PR  
XX (UYVA ) UNIV YALE.  
PA  
XX Lifton RP, Simon DB;  
PI  
XX WPI; 1998-388029/33.  
DR

Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2Cl cotransporter genes - useful for developing products for the diagnosis and treatment of ion transport disorders, e.g. Gitelman's Syndrome or Bartter's Syndrome

Example 2; Page 65; 105pp; English.

Primers hNKCC2ex19 forward and reverse (see AAV45715 and AAV45716, respectively) are designed to amplify exon 19 of the human NKCC2 gene (see AAV40562) that codes for Na-K-2Cl cotransporter NKCC2 protein (see AAW29683). Both primers are located within introns of the gene. 27 Sets of specific primers (see AAV45677-V45730) were used for SSCP analysis of NKCC2. Amplified products were analysed for molecular variants by electrophoresis, and identified variants CC were sequenced. Complete linkage of Bartter's syndrome with NKCC2 CC was demonstrated. Identification of the molecular basis of CC Bartter's syndrome allows for the genetic diagnosis of this CC disorder. The invention provides products and methods useful for CC diagnosis and treatment of Bartter's syndrome and other ion CC transport disorders.

XX  
SQ Sequence 23 BP; 8 A; 4 C; 5 G; 6 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 23;  
Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9  
DB 10 CTAGTATTA 2

RESULT 7  
AAZ61141  
ID AAZ61141 standard; DNA; 28 BP.  
XX  
AC AAZ61141;  
XX  
DT 30-MAY-2000 (first entry)  
XX  
DE Reverse PCR primer used to amplify DNA encoding the Shiga toxin.  
XX  
XX Chemokine receptor; ligand; inflammatory response; immune effector cell;  
KW secondary tissue damage; central nervous system injury; Shiga;  
KW CNS inflammatory disease; neurodegenerative disorder; heart disease;  
KW inflammatory eye disease; inflammatory bowel disease; PCR primer;  
KW inflammatory joint disease; inflammatory kidney; renal disease;  
KW inflammatory lung disease; inflammatory nasal disease; thyroiditis;  
KW inflammatory thyroid disease; cytokine-regulated cancer; ss.  
XX  
OS Shigella dysenteriae.  
XX  
XX WO200004926-A2.  
XX  
XX 03-FEB-2000.  
PD  
XX 21-JUL-1999; 99WO-CA00659.  
PF  
XX 22-JUL-1998; 98US-0120523.  
PR  
XX (OSPR-) OSPREY PHARM LTD.  
PA  
XX McDonald JR, Coggins PJ;  
PI  
XX WPI; 2000-182542/16.  
DR

A new therapeutic agent comprising a conjugate for treating secondary tissue damage and other disease conditions like Alzheimer's disease, stroke, Parkinson's disease and atherosclerosis

Example 1; Page 135; 204pp; English.

PCR primers AAZ61140-42 were used to amplify nucleic acids encoding the Shiga toxin. The toxin can be incorporated into the conjugates of the invention. The specification describes a conjugate, comprising a targeted agent and a chemokine receptor ligand. The conjugate binds to a chemokine receptor resulting in internalisation of the CC targeted agent in cells bearing the receptor. The conjugates are CC used for formulating a medicament or for treating disorders associated CC with inflammatory responses resulting from activation, proliferation CC and migration of immune effector cells. The disorders or disease states CC comprise secondary tissue damage such as central nervous system (CNS) CC injury, CNS inflammatory diseases, neurodegenerative disorders, heart CC disease, inflammatory eye diseases, inflammatory bowel diseases, CC inflammatory joint diseases, inflammatory kidney or renal diseases, CC inflammatory lung diseases, inflammatory nasal diseases, inflammatory CC thyroid disease such as thyroiditis, or cytokine-regulated cancers.

XX  
SQ Sequence 28 BP; 6 A; 7 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9  
DB 11 ctagtattata 19

```

RESULT 8
AAZ87693
ID AAX87693 standard; DNA; 30 BP.
XX
AC AAX87693;
XX
DT 26-OCT-1999 (first entry)
XX
DE Human TPST-1 cDNA bottom strand PCR primer.
XX
TPST-1; tyrosylprotein sulfotransferase; tyrosine O-sulfation;
KW post-translation modification; human; PCR; primer; ss.
KW
XX
OS Synthetic.
OS Homo sapiens.
OS
XX
PN WO9938980-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01756.
XX
PR 09-SEP-1998; 98US-0150141.
XX
PR 29-JAN-1998; 98US-0072994.
XX
PR 09-SEP-1998; 98US-0150133.
XX
PA (OKLA ) UNIV OKLAHOMA STATE.
XX
PI Moore KL;
XX
DR WPI; 1999-494094/41.
XX
PT Human, murine and Caenorhabditis elegans tyrosylprotein
PT sulfotransferases, useful for post-translational tyrosine sulfation
XX
XX Example; Page 39; 123pp; English.
XX
This bottom strand primer was used with a top strand primer (see
CC AAX87692) in the PCR amplification of the human tyrosylprotein
CC sulfotransferase (TPST-1) coding sequence (see AAX87686) using EST
CC clone #116978 as template. The primers introduce a 5' BamHI site
CC and a 3' SpeI site. The amplified cDNA was ligated into vector
CC pCDNA3.1(+), and TPST-1 (see AAY06623) was expressed as a fusion
CC protein containing an N-terminal epitope for HPC4 (see AAY06630)
CC in mammalian cells. TPST-1 is useful for the post-translational
CC tyrosine O-sulfation of proteins and peptides.
XX
XX Sequence 30 BP; 5 A; 9 C; 6 G; 10 T; 0 other;
XX
Query Match 100.0%; Score 9; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctagtattta 9
Db 4 ctagtattta 12
XX
RESULT 9
AAZ99717
ID AAX99717 standard; DNA; 30 BP.
XX
AC AAX99717;
XX
DT 12-JUL-2000 (first entry)
XX
DE PCR primer for cDNA encoding human tyrosylprotein sulfotransferase 1.
XX
Human; tyrosylprotein sulfotransferase 1; TPST-1; tyrosine O-sulfation;
KW P-selectin glycoprotein ligand-1; PSGL-1; anti-inflammatory;
KW neutrophil binding; PCR primer; ss.
XX
XX

```

---

```

OS Homo sapiens.
XX
PN WO200014250-A1.
XX
PD 16-MAR-2000.
XX
PF 23-JUL-1999; 99WO-US16750.
XX
PR 09-SEP-1998; 98US-0150133.
XX
PR 09-SEP-1998; 98US-0150141.
XX
PR 28-JAN-1999; 99WO-US01756.
XX
PA (OKLA ) UNIV OKLAHOMA STATE.
XX
PI Moore KL;
XX
DR WPI; 2000-271057/23.
XX
PT New human or murine tyrosylprotein sulfotransferase, used for in vitro
PT or in vivo sulfation of proteins and to screen for anti-inflammatory
PT agents.
XX
XX Example; Page 40; 141pp; English.
XX
PCR primers AAZ99716-17 were used to amplify cDNA encoding human
CC tyrosylprotein sulfotransferase 1 (TPST-1) polypeptide. TPST-1 catalyses
CC the transfer of sulfate from 3'-phosphoadenosine-5'-phosphosulfate to
CC tyrosine residues (tyrosine O-sulfation) in proteins. TPST-1
CC polypeptides are used for in vitro or in vivo sulfation of proteins or
CC peptides, particularly P-selectin glycoprotein ligand-1 (PSGL-1) and its
CC functional fragments. They are also used to raise specific antibodies,
CC and are useful as immunoassay reagents, and to identify specific
CC inhibitors. These inhibitors are potentially useful as anti-inflammatory
CC agents, by inhibiting PSGL-1 mediated binding of neutrophils to
CC endothelial cells. The TPST-1 polynucleotide is useful for identifying
CC polymorphisms that may be associated with a disease phenotype.
XX
XX Sequence 30 BP; 5 A; 9 C; 6 G; 10 T; 0 other;
XX
Query Match 100.0%; Score 9; DB 21; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctagtattta 9
Db 4 ctagtattta 12
XX
RESULT 10
AAQ56057/c
ID AAQ56057 standard; DNA; 31 BP.
XX
AC AAQ56057;
XX
DT 12-AUG-1994 (first entry)
XX
DE Coconut Foliar Decay Virus promoter stem-loop.
XX
KW Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;
KW phloem-specific; stem-loop structure; transgenic plant; ds.
XX
OS Coconut Foliar Decay Virus.
XX
PH Key Location/Qualifiers
FT stem_loop 1..31
FT /*tag= a
FT /function= promoter
FT /note= "loop has homology to geminivirus sequence"
XX
XX DE4306832-C.
XX
XX 24-FEB-1994.
XX

```

XX PF 04-MAR-1993; 93DE-4306832.  
XX PR 04-MAR-1993; 93DE-4306832.  
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX PI Becker D, Randles JW, Rohde W, Salamini F;  
XX DR WPI; 1994-058406/08.  
XX PT Use of coconut foliar decay virus DNA as promoter - for  
XX PT tissue-specific gene expression in transgenic plants  
XX PS Claim 1; Fig 2; 8pp; German.  
XX CC A DNA fragment from the CFV genome can be used as a phloem-specific  
XX CC promoter in the construction of transgenic plants. The promoter is  
XX CC strong; it has 30-50% of the activity of the CaMV 35S promoter in  
XX CC tobacco plants.  
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;  
  
Query Match 100.0%; Score 9; DB 15; Length 31;  
Best Local Similarity 100.0%; Pred. No. 3.6e+03; Indels 0; Caps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
  
QY 1 ctagtattata 9  
DB 20 CTAGTATTATA 12  
  
RESULT 11  
AAX02688/C  
ID AAX02688 standard; DNA; 31 BP.  
XX AC AAX02688;  
XX DT 10-MAY-1999 (first entry)  
XX DE Gemini virus DNA fragment stem loop.  
XX KW Stem loop; coconut foliar decay virus; CFV; bacterial promoter;  
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.  
XX OS Gemini virus.  
XX FH Location/Qualifiers  
FT stem\_loop 1..31 /\*tag= a  
FT misc\_binding 1..11 /\*tag= b  
FT misc\_binding 21..31 /\*tag= c  
FT /\*note= "Region binds to nucleotides 21 to 31"  
FT /\*tag= c  
FT /\*note= "Region binds to nucleotides 1 to 11"  
XX DE19730502-A1.  
XX PN 21-JAN-1999.  
XX PD 16-JUL-1997; 97DE-1030502.  
XX PF 16-JUL-1997; 97DE-1030502.  
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;  
XX PI WPI; 1999-096863/09.  
XX DR Coconut foliar decay virus promoters - for gene expression in  
XX PT

PT bacteria and yeasts  
XX Disclosure; Fig 2; 14pp; German.  
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment  
XX CC that includes the stem-loop structure of CFV DNA but lacks the  
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The  
XX CC new CFV DNA fragment is useful as a bacterial or yeast promoter, as a  
XX CC promoter for tissue-specific (especially phloem-specific) gene expression  
XX CC in plants and for production of chimeric constructs for transient or  
XX CC stable expression. Certain fragments of CFV DNA have stronger promoter  
XX CC activity in E. coli than the CaMV 35S promoter.  
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;  
  
Query Match 100.0%; Score 9; DB 20; Length 31;  
Best Local Similarity 100.0%; Pred. No. 3.6e+03; Indels 0; Caps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
  
QY 1 ctagtattata 9  
DB 20 CTAGTATTATA 12  
  
RESULT 12  
AAX02686/C  
ID AAX02686 standard; DNA; 31 BP.  
XX AC AAX02686;  
XX DT 10-MAY-1999 (first entry)  
XX DE Gemini virus DNA fragment stem loop.  
XX KW Stem loop; coconut foliar decay virus; CFV; bacterial promoter;  
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.  
XX OS Gemini virus.  
XX FH Location/Qualifiers  
FT stem\_loop 1..31 /\*tag= a  
FT misc\_binding 1..11 /\*tag= b  
FT misc\_binding 21..31 /\*tag= c  
FT /\*note= "Region binds to nucleotides 21 to 31"  
FT /\*tag= c  
FT /\*note= "Region binds to nucleotides 1 to 11"  
XX DE19730535-A1.  
XX PN 21-JAN-1999.  
XX PD 16-JUL-1997; 97DE-1030535.  
XX PF 16-JUL-1997; 97DE-1030535.  
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;  
XX PI WPI; 1999-096867/09.  
XX DR Coconut foliar decay virus promoters - for gene expression in  
XX PT plants, bacteria and yeasts  
XX PS Disclosure; Fig 2; 12pp; German.  
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment  
XX CC that includes the stem-loop structure of CFV DNA but lacks the  
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The  
XX CC new CFV DNA fragment is useful as a bacterial or yeast promoter, as a

CC promoter for tissue-specific (especially phloem-specific) gene expression  
 CC in plants and for production of chimeric constructs for transient or  
 CC stable expression. Certain fragments of CPDV DNA have stronger promoter  
 CC activity in *E. coli* than the CaMV 35S promoter.  
 XX  
 SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9  
 |||||  
 Db 20 CTAGTATTATA 12

## RESULT 13

AAT90052  
 ID AAT90052 standard; DNA; 32 BP.

XX  
 AC AAT90052;

XX  
 DT 27-MAR-1998 (first entry)

XX  
 DE Primer for chicken urokinase.cDNA.

XX  
 KW Low molecular weight; chicken; urokinase; reporter; screening;  
 PCR primer; ss.

XX  
 OS Gallus domesticus.

XX  
 PN WO9734144-AL.

XX  
 PD 18-SEP-1997.

XX  
 PF 11-MAR-1997; 97WO-EP01219.

XX  
 PR 20-DEC-1996; 96DE-4054616.

XX  
 PR 14-MAR-1996; 96DE-4011780.

XX  
 PA (SCHD ) SCHERING AG.

XX  
 PI Langer G, Schleuning W., Toschi L;

XX  
 DR WPI; 1997-470977/43.

XX  
 PT Determining effect of test substance on cultured cells - by  
 PT transfecting cells with plasmid having gene for low molecular weight  
 PT hen urokinase as reporter, which can be detected by plasminogen  
 PT activation and colour forming or fluorochrome substrate reaction  
 XX

PS Example 1; Page 16; 32pp; German.

XX  
 CC The present sequence was used in the development of a method for  
 CC determining the effect of test substance in a cell culture system.  
 CC The method comprises preparing a plasmid having an expression  
 CC cassette including the low molecular weight gene for chicken  
 CC urokinase as a reporter, transfecting eukaryotic cells  
 CC with the plasmid, inducing the cells with the test substance,  
 CC harvesting the cell supernatant and detecting urokinase by adding  
 CC chicken plasminogen (this addition precisely defines the start of  
 CC the reaction), so that this is converted to plasmin by the  
 CC urokinase, and detecting the plasmin in the supernatant with a  
 CC highly sensitive chromogenic or fluorochromic substrate. The method  
 CC can be used to screen new compounds, non-steroidal glucocorticoids  
 CC or anti-glucocorticoids, determine receptors present in cells,  
 CC particularly steroid hormone receptors and investigate promoter  
 CC dependent modulated transcription processes. The method does not  
 CC require radioisotopes or extraction of reporter gene products, is  
 CC sensitive, non-destructive and can be performed in microtitre  
 CC plates. Urokinase is small (so easily separated from endogenous  
 CC proteases by gel electrophoresis), and unlike other reporter gene

CC products is stable in culture supernatant, permitting direct  
 CC measurement. The method can be calibrated using a precise number of  
 CC transfected cells, avoiding difficulties of calibration based on  
 CC enzymatic activity. The urokinase/plasminogen reaction represents a  
 CC long lived amplification system, which is well characterised.  
 XX

SQ Sequence 32 BP; 5 A; 7 C; 11 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 18; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9  
 |||||  
 Db 12 ctagtattata 20

## RESULT 14

AAT77100

ID AAT77100 standard; DNA; 35 BP.

XX  
 AC AAT77100;

XX  
 DT 20-JAN-1998 (first entry)

XX  
 DE Pisum sativum plastocyanin promoter sub-sequence.

XX  
 KW plastocyanin; promoter; enhancer; reporter gene; pea;  
 transgenic plant; ds.

XX  
 OS Pisum sativum.

XX  
 FH Key Location/Qualifiers  
 FT enhancer 1..31

FT misc\_feature 1..4

FT /tag= a

FT /tag= b

FT /label= sticky\_end

FT /note= "The 5' end of complementary strand

FT overhangs the 3' end of this strand by the

FT sequence 5'-GATC-3'."

XX  
 PN WO9720056-A2.

XX  
 PD 05-JUN-1997.

XX  
 PF 26-NOV-1996; 96WO-GB02910.

XX  
 PR 29-NOV-1995; 95GB-0024350.

XX  
 PA (ADTE-) ADVANCED TECHNOLOGIES CAMBRIDGE.

XX  
 PI Gray JC, Sandhu JS, Webster CI;

XX  
 DR WPI; 1997-310610/28.

XX  
 PT Enhancer from pea plastocyanin promoter region - used to produce

XX  
 PT transgenic plants having increased expression of a desired gene

XX  
 PS Claim 11; Figure 7; 46pp; English.

XX  
 CC This sequence represents the pea plastocyanin promoter enhancer  
 CC which is a sub-sequence of AAT77099. This enhancer has been used to  
 CC increase the expression of the GUS reporter gene in tobacco where the  
 CC GUS gene was under the control of the PetE promoter and 35S CaMV  
 CC promoter. The enhancer was also used joined to the -330 to +1 region of  
 CC the patatin promoter PS20 which was joined to the GUS reporter and  
 CC transformed into potato using A.tumefaciens. Expression of the gene  
 CC promoter of one or more genes of a plant increases in green or non-green  
 CC tissues, especially roots, tubers, seeds, flowers or leaves and  
 CC transformed plants may be dicotyledonous or monocotyledonous. This  
 CC enhancer may operate both in normal and reverse directions and can be

Thu Oct 25 13:08:57 2001

CC attached to the promoter or terminator of the gene to be expressed.

XX  
SQ Sequence 35 BP; 17 A; 4 C; 2 G; 12 T; 0 other;  
Query Match 100.0%; Score 9; DB 18; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctagtattata 9  
|||||  
Db 11 ctagtattata 19

RESULT 15  
AAA39333  
ID AAA39333 standard; DNA; 35 BP.  
XX  
AC AAA39333;  
XX 11-SEP-2000 (first entry)  
XX Z. paucivorans 16S rRNA and 23S rRNA gene spacer region SEQ ID NO:8.  
XX 16S rRNA; 23S rRNA; spacer region; detection; microbe; beer;  
KW brewery; ds.  
XX Zymophilus paucivorans.  
XX JP2000106881-A.  
PN 18-APR-2000.  
PD 08-OCT-1998; 98JP-0286697.  
PF 08-OCT-1998; 98JP-0286697.  
PR (ASAK) ASAHI BREWERIES LTD.  
XX WPI; 2000-353477/31.  
DR A gene sequence spacer region between a gene encoding 16S rRNA and a  
PT gene encoding 23S rRNA, useful for detection of microbes -  
XX Claim 8; Page 2; 18pp; Japanese.  
PS The present invention describes gene sequences of spacer region between  
XX a gene encoding 16S rRNA and a gene encoding 23S rRNA for the detection  
CC of microbes. Also described is a method for the detection of microbes  
CC in which an oligonucleotide prepared from a spacer region as described  
CC above is made to function as a primer for nucleic acid synthesis and  
CC treated by gene amplification. The method is used for the detection of  
CC microbes found in beer breweries. The present sequence represents a  
CC specifically claimed Zymophilus paucivorans 16S rRNA and 23S rRNA spacer  
CC region DNA sequence, from the present invention.  
XX  
SQ Sequence 35 BP; 11 A; 3 C; 7 G; 14 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3.6e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctagtattata 9  
|||||  
Db 7 ctagtattata 15

Search completed: October 24, 2001, 10:07:28  
Job time: 633 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:31 ; Search time 180.6 seconds  
(without alignments)  
9.434 Million cell updates/sec

Title: US-09-462-955-5  
Perfect score: 9  
Sequence: 1 ctgattatta 9

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Query | Score | Match | Length | DB                | ID | Description        |
|------------|-------|-------|-------|--------|-------------------|----|--------------------|
| 1          | 9     | 100.0 | 12    | 2      | US-08-418-071-15  |    | Sequence 15, Appl  |
| 2          | 9     | 100.0 | 30    | 3      | US-09-150-133-33  |    | Sequence 33, Appl  |
| 3          | 9     | 100.0 | 30    | 3      | US-09-150-141-33  |    | Sequence 33, Appl  |
| 4          | 9     | 100.0 | 30    | 4      | US-09-374-493-33  |    | Sequence 33, Appl  |
| 5          | 9     | 100.0 | 30    | 4      | US-09-374-824-33  |    | Sequence 33, Appl  |
| 6          | 9     | 100.0 | 30    | 4      | US-09-374-492-33  |    | Sequence 33, Appl  |
| 7          | 9     | 100.0 | 38    | 3      | US-08-817-926-41  |    | Sequence 41, Appl  |
| 8          | 9     | 100.0 | 41    | 3      | US-08-491-954-106 |    | Sequence 106, Appl |
| 9          | 9     | 100.0 | 62    | 3      | US-08-492-076-9   |    | Sequence 9, Appl   |
| 10         | 9     | 100.0 | 78    | 2      | US-08-680-326-51  |    | Sequence 51, Appl  |
| 11         | 9     | 100.0 | 101   | 1      | US-07-755-859B-14 |    | Sequence 14, Appl  |
| 12         | 9     | 100.0 | 101   | 1      | US-08-457-616-14  |    | Sequence 14, Appl  |
| 13         | 9     | 100.0 | 299   | 2      | US-08-637-759B-31 |    | Sequence 31, Appl  |
| 14         | 9     | 100.0 | 299   | 2      | US-08-871-355A-31 |    | Sequence 31, Appl  |
| 15         | 9     | 100.0 | 352   | 6      | 5242821-11        |    | Patent No. 5242821 |
| 16         | 9     | 100.0 | 454   | 1      | US-08-846-134-2   |    | Sequence 2, Appl   |
| 17         | 9     | 100.0 | 492   | 1      | US-08-470-720-3   |    | Sequence 3, Appl   |
| 18         | 9     | 100.0 | 573   | 1      | US-08-709-912-18  |    | Sequence 18, Appl  |
| 19         | 9     | 100.0 | 573   | 2      | US-09-047-370-18  |    | Sequence 18, Appl  |
| 20         | 9     | 100.0 | 599   | 3      | US-08-705-875A-2  |    | Sequence 2, Appl   |
| 21         | 9     | 100.0 | 600   | 5      | PCT-US91-02766-24 |    | Sequence 24, Appl  |
| 22         | 9     | 100.0 | 628   | 2      | US-08-874-186-45  |    | Sequence 24, Appl  |
| 23         | 9     | 100.0 | 660   | 3      | US-09-284-782-15  |    | Sequence 45, Appl  |
| 24         | 9     | 100.0 | 663   | 4      | US-08-998-416-187 |    | Sequence 15, Appl  |
| 25         | 9     | 100.0 | 706   | 3      | US-08-654-025-4   |    | Sequence 187, Appl |
| 26         | 9     | 100.0 | 713   | 1      | US-08-234-939-9   |    | Sequence 9, Appl   |
| 27         | 9     | 100.0 | 713   | 1      | US-08-558-865-9   |    | Sequence 9, Appl   |

|    |   |       |     |   |                    |                    |
|----|---|-------|-----|---|--------------------|--------------------|
| 28 | 9 | 100.0 | 717 | 4 | US-08-998-416-849  | Sequence 849, App  |
| 29 | 9 | 100.0 | 719 | 4 | US-08-998-416-1138 | Sequence 1138, App |
| 30 | 9 | 100.0 | 720 | 4 | US-08-998-416-628  | Sequence 628, App  |
| 31 | 9 | 100.0 | 746 | 4 | US-08-991-789A-261 | Sequence 261, App  |
| 32 | 9 | 100.0 | 836 | 1 | US-09-077-675A-11  | Sequence 11, Appl  |
| 33 | 9 | 100.0 | 837 | 1 | US-08-832-883-56   | Sequence 56, Appl  |
| 34 | 9 | 100.0 | 837 | 2 | US-08-832-877-56   | Sequence 56, Appl  |
| 35 | 9 | 100.0 | 839 | 3 | US-08-817-926-50   | Sequence 50, Appl  |
| 36 | 9 | 100.0 | 856 | 4 | US-08-998-416-289  | Sequence 289, App  |
| 37 | 9 | 100.0 | 890 | 1 | US-08-234-939-8    | Sequence 8, Appl   |
| 38 | 9 | 100.0 | 890 | 1 | US-08-558-865-8    | Sequence 8, Appl   |
| 39 | 9 | 100.0 | 943 | 3 | US-08-705-875A-3   | Sequence 3, Appl   |
| 40 | 9 | 100.0 | 967 | 4 | US-08-960-780-47   | Sequence 47, Appl  |
| 41 | 9 | 100.0 | 967 | 4 | US-09-073-898-47   | Sequence 47, Appl  |
| 42 | 9 | 100.0 | 968 | 3 | US-08-705-875A-9   | Sequence 9, Appl   |
| 43 | 9 | 100.0 | 972 | 4 | US-08-960-780-48   | Sequence 48, Appl  |
| 44 | 9 | 100.0 | 972 | 4 | US-09-073-898-48   | Sequence 48, Appl  |
| 45 | 9 | 100.0 | 975 | 3 | US-09-015-754-1    | Sequence 1, Appl   |

## ALIGNMENTS

RESULT 1  
US-08-418-071-15  
; Sequence 15, Application US/08418071  
; Patent No. 5846705  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Rey-Yuh  
; APPLICANT: You, Li-Ru  
; APPLICANT: Soong, Tai-Seng  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA  
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/418,071  
; FILING DATE: 06-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: DCB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: circular  
; MOLECULE TYPE: other nucleic acid  
; ANTI-SENSE: NO  
US-08-418-071-15

Query Match 100.0%; Score 9; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

us-09-462-955-5.rni

Thu Oct 25 13:08:59 2001

```
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-493-33

Query Match      100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
Db 4 ctagtattta 12

RESULT 5
US-09-374-824-33
; Sequence 33, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-824-33

Query Match      100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
Db 4 ctagtattta 12

RESULT 6
US-09-374-492-33
; Sequence 33, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
```

```
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-150-133-33

Query Match      100.0%; Score 9; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
Db 4 ctagtattta 12

RESULT 3
US-09-150-141-33
; Sequence 33, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-150-141-33

Query Match      100.0%; Score 9; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
Db 4 ctagtattta 12

RESULT 4
US-09-374-493-33
; Sequence 33, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
```

FILE REFERENCE: 5820.545  
CURRENT APPLICATION NUMBER: US/09/374,492  
CURRENT FILING DATE: 1999-08-13  
EARLIER APPLICATION NUMBER: 09/150,141  
EARLIER FILING DATE: 1998-09-09  
EARLIER APPLICATION NUMBER: 60/072,994  
EARLIER FILING DATE: 1998-01-29  
EARLIER APPLICATION NUMBER: PCT/US99/16750  
EARLIER FILING DATE: 1999-07-23  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: Wordperfect 8.0 (saved in ASCII format)  
SEQ ID NO 33  
LENGTH: 30  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: primer  
US-09-374-492-33

Query Match 100.0%; Score 9; DB 4; Length 30;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9  
Db 4 ctagtattta 12

RESULT 7  
US-0817-926-41/c  
Sequence 41, Application US/08817926  
Patent No. 6001590  
GENERAL INFORMATION:  
APPLICANT: Komeda, Toshihiro  
APPLICANT: Suda, Hisako  
APPLICANT: Tamai, Yukio  
APPLICANT: Iwamatsu, Akihiro  
APPLICANT: Kato, No. 6001590uo  
APPLICANT: Sakai, Yasuyoshi  
TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINII  
TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC-compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,926  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/02597  
FILING DATE: 12-SEP-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 234133/1995  
FILING DATE: 12-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 42536/1996  
FILING DATE: 29-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 081356/0112  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic DNA"  
US-08-817-926-41

Query Match 100.0%; Score 9; DB 3; Length 38;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9  
Db 18 CTAGTATTA 10

RESULT 8  
US-08-491-954-106  
Sequence 106, Application US/08491954  
Patent No. 6096321  
GENERAL INFORMATION:  
APPLICANT: Girardeau, Jean-Pierre  
APPLICANT: Martin, Christine  
APPLICANT: Mechin, Marie-Claire  
APPLICANT: Der Vartanian, Maurice  
APPLICANT: Bousquet, Francois  
TITLE OF INVENTION: SUB-UNIT OF CS31A PROTEIN CAPSULE  
TITLE OF INVENTION: MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A  
TITLE OF INVENTION: PROTEIN CAPSULE INCLUDING SUCH A SUB-UNIT, AND  
TITLE OF INVENTION: MICROORGANISMS WHOSE OUTER MEMBRANE CARRIES SUCH  
TITLE OF INVENTION: SUB-UNITS, AND PROCEDURE FOR OBTAINING AND UTILIZING SUCH  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WEISER & ASSOCIATES  
STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/491,954  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/01281  
FILING DATE: 21-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989.6264P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 41 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna

FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..41  
US-08-491-954-106

Query Match 100.0%; Score 9; DB 3; Length 41;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtatta 9  
| | | | |  
Db 1 CTAGTATTA 9

## RESULT 9

US-08-492-076-9  
; Sequence 9, Application US/08492076A  
; Patent No. 6060064  
; GENERAL INFORMATION:  
; APPLICANT: Adams, Sally E.  
; APPLICANT: Burus, Nigel R.  
; APPLICANT: Richardson, Simon M.  
; TITLE OF INVENTION: No. 6060064e1 Proteinaceous Particles  
; FILE REFERENCE: 10180.60968  
; CURRENT APPLICATION NUMBER: US/08/492.076A  
; EARLIER FILING DATE: 1995-06-28  
; EARLIER APPLICATION NUMBER: PCT/GB93/02656  
; EARLIER FILING DATE: 1993-12-24  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 62  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Partial V3  
; OTHER INFORMATION: loop of HIV-1 further comprising restriction  
; OTHER INFORMATION: enzyme sites.  
US-08-492-076-9

Query Match 100.0%; Score 9; DB 3; Length 62;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtatta 9  
| | | | |  
Db 1 ctagtatta 9

## RESULT 10

US-08-680-326-51  
; Sequence 51, Application US/08680326  
; Patent No. 5925733  
; GENERAL INFORMATION:  
; APPLICANT: ROSE, TIMOTHY M.  
; APPLICANT: BOSCH, MARINX  
; APPLICANT: STRAND, KURT  
; APPLICANT: TODARO, GEORGE J.  
; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES  
; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL  
; TITLE OF INVENTION: FIBROMATOSIS  
; NUMBER OF SEQUENCES: 152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FORSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680.326  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Schiff, J. Michael  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-680-326-51

Query Match 100.0%; Score 9; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtatta 9  
| | | | |  
Db 29 CTAGTATTA 37

## RESULT 11

US-07-795-859B-14/c  
; Sequence 14, Application US/07795859B  
; Patent No. 5422262  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Stefan  
; APPLICANT: Russell, David W.  
; TITLE OF INVENTION: Steroid 5 $\alpha$ -Reductases  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White and Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/795.859B  
; FILING DATE: 18-NOV-1991  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UTSD:260/PAR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 320-7200  
; TELEFAX: (512) 474-7677  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 101 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-07-795-859B-14

Query Match 100.0%; Score 9; DB 1; Length 101;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattat 9  
Db 26 CTAGTATTA 18

RESULT 12

US-08-457-616-14/c  
; Sequence 14, Application US/08457616  
; Patent No. 5679521  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Stefan  
; APPLICANT: Russell, David W.  
; TITLE OF INVENTION: Steroid 5 $\alpha$ -Reductases  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White and Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,616  
; FILING DATE: 01-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/795,859  
; FILING DATE: 18-NOV-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UTSD:260/PAR  
; TELEPHONE: (512) 320-7200  
; TELEFAX: (512) 474-7677  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 101 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-457-616-14

Query Match 100.0%; Score 9; DB 1; Length 101;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattat 9  
Db 26 CTAGTATTA 18

RESULT 13

US-08-637-759B-31/c  
; Sequence 31, Application US/08637759B  
; Patent No. 5876931  
; GENERAL INFORMATION:  
; APPLICANT: David William Holden  
; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/637,759B  
; FILING DATE: 03-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB95/02875  
; FILING DATE: 11-DEC-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: RPLS 101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 873-8794  
; TELEFAX: (404) 873-8795  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 299 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Partial sequence of Salmonella typhimurium  
; ORGANISM: virulence gene  
US-08-637-759B-31

Query Match 100.0%; Score 9; DB 2; Length 299;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattat 9  
Db 52 CTAGTATTA 44

RESULT 14

US-08-871-355A-31/c  
; Sequence 31, Application US/08871355A  
; Patent No. 6015669  
; GENERAL INFORMATION:  
; APPLICANT: David William Holden  
; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

Thu Oct 25 13:08:59 2001

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
; US-08-871-355A-31

```

```

Query Match      100.0%; Score 9; DB 3; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ctagtattata 9
    |||
DB 52 CTAGTATTA 44

```

```

RESULT 15
5242821-11
; Patent No. 5242821
; APPLICANT: PALVA, LIKKA;SIBAKHOV, MERVI
; TITLE OF INVENTION: LACTOCOCCUS PROMOTER AND SIGNAL
; SEQUENCES FOR EXPRESSION IN BACTERIA
; NUMBER OF SEQUENCES: 27
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/377,450
; FILING DATE: 10-JUL-1989
; SEQ ID NO:11:
; LENGTH: 352
5242821-11

```

```

Query Match      100.0%; Score 9; DB 6; Length 352;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ctagtattata 9
    |||
DB 131 ctagtattata 139

```

```

Search completed: October 24, 2001, 10:00:31
Job time: 216 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:59 ; Search time 5479.82 Seconds  
(without alignments)  
15.525 Million cell updates/sec

Title: US-09-462-955-5  
Perfect score: 9  
Sequence: 1 ctagtattta 9

Scoring table: IDENTITY\_NUC-  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est33:\*  
111: gb\_est34:\*  
112: gb\_est35:\*  
113: gb\_est36:\*  
114: gb\_est37:\*  
115: gb\_est38:\*  
116: gb\_est39:\*

Thu Oct 25 13:09:00 2001

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description        |
|------------|-------|---------|--------------|-------|--------------------|
| 1          | 9     | 100.0   | 40           | 167   | BE386120 601277233 |
| 2          | 9     | 100.0   | 55           | 249   | AZ769245 1M0569010 |
| 3          | 9     | 100.0   | 57           | 243   | AF097417 AF097417  |
| 4          | 9     | 100.0   | 57           | 243   | AZ434302 1M0220E09 |
| 5          | 9     | 100.0   | 58           | 10    | AA637860 VR29f01.r |
| 6          | 9     | 100.0   | 59           | 113   | AW215812 uo99e06.x |
| 7          | 9     | 100.0   | 59           | 113   | AW215820 uo99f06.x |
| 8          | 9     | 100.0   | 62           | 258   | TA82F100 T. brucei |
| 9          | 9     | 100.0   | 66           | 250   | AZ808515 2M0072A07 |
| 10         | 9     | 100.0   | 67           | 24    | A1761414 w65c06.x  |
| 11         | 9     | 100.0   | 71           | 4     | AA250291 m261g02.r |
| 12         | 9     | 100.0   | 71           | 123   | AA989354 u25p07.y  |
| 13         | 9     | 100.0   | 71           | 138   | BE627067 u15f06.y  |
| 14         | 9     | 100.0   | 71           | 138   | BE630111 u15f06.x  |
| 15         | 9     | 100.0   | 71           | 138   | BE688187 u43b06.x  |
| 16         | 9     | 100.0   | 71           | 138   | BE692587 u43b06.y  |
| 17         | 9     | 100.0   | 71           | 141   | BE849664 u43b06.y  |
| 18         | 9     | 100.0   | 73           | 167   | BE448984 u50h06.y  |
| 19         | 9     | 100.0   | 75           | 2     | AA079298 zm97g12.s |
| 20         | 9     | 100.0   | 78           | 113   | AW226900 u63c01.y  |
| 21         | 9     | 100.0   | 80           | 138   | BE627009 u08h09.y  |
| 22         | 9     | 100.0   | 84           | 2     | AA079297 zm97g12.r |
| 23         | 9     | 100.0   | 84           | 106   | AU012912 AU012912  |
| 24         | 9     | 100.0   | 85           | 13    | AA937702 OJ04g12.s |
| 25         | 9     | 100.0   | 87           | 111   | AA760419 v76h02.r  |
| 26         | 9     | 100.0   | 87           | 249   | AZ783472 2M0025B18 |
| 27         | 9     | 100.0   | 90           | 155   | C01576 HMG5000858  |
| 28         | 9     | 100.0   | 93           | 159   | N28044 EST000258 S |
| 29         | 9     | 100.0   | 94           | 20    | A1465178 mv98a02.x |
| 30         | 9     | 100.0   | 94           | 241   | AZ309606 1M0016B06 |
| 31         | 9     | 100.0   | 94           | 247   | AZ660200 1M0538I08 |
| 32         | 9     | 100.0   | 95           | 31    | AV562434 AV562434  |
| 33         | 9     | 100.0   | 97           | 7     | AA419984 VF38h02.r |
| 34         | 9     | 100.0   | 99           | 107   | AU014567 AU014567  |
| 35         | 9     | 100.0   | 100          | 10    | AA689650 v50g11.r  |
| 36         | 9     | 100.0   | 100          | 24    | A1718783 as58h08.x |
| 37         | 9     | 100.0   | 101          | 164   | BE152001 QV1-HT031 |
| 38         | 9     | 100.0   | 101          | 164   | BE179567 IL3-HT061 |
| 39         | 9     | 100.0   | 102          | 186   | BE365112 P11.24.B0 |
| 40         | 9     | 100.0   | 102          | 189   | T61718 yb92e07.r1  |
| 41         | 9     | 100.0   | 102          | 237   | AZ059959 RPCI-23-4 |
| 42         | 9     | 100.0   | 104          | 8     | AA518125 v123g11.r |
| 43         | 9     | 100.0   | 104          | 121   | AW810470 MR4-ST012 |
| 44         | 9     | 100.0   | 104          | 164   | BE145670 ILO-HT020 |
| 45         | 9     | 100.0   | 105          | 1     | AA014683 mh09g05.r |

## ALIGNMENTS

| Result No. | Score | % Match | Query Length | DB ID | Description        |
|------------|-------|---------|--------------|-------|--------------------|
| 1          | 9     | 100.0   | 40           | 167   | BE386120 601277233 |
| 2          | 9     | 100.0   | 55           | 249   | AZ769245 1M0569010 |
| 3          | 9     | 100.0   | 57           | 243   | AF097417 AF097417  |
| 4          | 9     | 100.0   | 57           | 243   | AZ434302 1M0220E09 |
| 5          | 9     | 100.0   | 58           | 10    | AA637860 VR29f01.r |
| 6          | 9     | 100.0   | 59           | 113   | AW215812 uo99e06.x |
| 7          | 9     | 100.0   | 59           | 113   | AW215820 uo99f06.x |
| 8          | 9     | 100.0   | 62           | 258   | TA82F100 T. brucei |
| 9          | 9     | 100.0   | 66           | 250   | AZ808515 2M0072A07 |
| 10         | 9     | 100.0   | 67           | 24    | A1761414 w65c06.x  |
| 11         | 9     | 100.0   | 71           | 4     | AA250291 m261g02.r |
| 12         | 9     | 100.0   | 71           | 123   | AA989354 u25p07.y  |
| 13         | 9     | 100.0   | 71           | 138   | BE627067 u15f06.y  |
| 14         | 9     | 100.0   | 71           | 138   | BE630111 u15f06.x  |
| 15         | 9     | 100.0   | 71           | 138   | BE688187 u43b06.x  |
| 16         | 9     | 100.0   | 71           | 138   | BE692587 u43b06.y  |
| 17         | 9     | 100.0   | 71           | 141   | BE849664 u43b06.y  |
| 18         | 9     | 100.0   | 73           | 167   | BE448984 u50h06.y  |
| 19         | 9     | 100.0   | 75           | 2     | AA079298 zm97g12.s |
| 20         | 9     | 100.0   | 78           | 113   | AW226900 u63c01.y  |
| 21         | 9     | 100.0   | 80           | 138   | BE627009 u08h09.y  |
| 22         | 9     | 100.0   | 84           | 2     | AA079297 zm97g12.r |
| 23         | 9     | 100.0   | 84           | 106   | AU012912 AU012912  |
| 24         | 9     | 100.0   | 85           | 13    | AA937702 OJ04g12.s |
| 25         | 9     | 100.0   | 87           | 111   | AA760419 v76h02.r  |
| 26         | 9     | 100.0   | 87           | 249   | AZ783472 2M0025B18 |
| 27         | 9     | 100.0   | 90           | 155   | C01576 HMG5000858  |
| 28         | 9     | 100.0   | 93           | 159   | N28044 EST000258 S |
| 29         | 9     | 100.0   | 94           | 20    | A1465178 mv98a02.x |
| 30         | 9     | 100.0   | 94           | 241   | AZ309606 1M0016B06 |
| 31         | 9     | 100.0   | 94           | 247   | AZ660200 1M0538I08 |
| 32         | 9     | 100.0   | 95           | 31    | AV562434 AV562434  |
| 33         | 9     | 100.0   | 97           | 7     | AA419984 VF38h02.r |
| 34         | 9     | 100.0   | 99           | 107   | AU014567 AU014567  |
| 35         | 9     | 100.0   | 100          | 10    | AA689650 v50g11.r  |
| 36         | 9     | 100.0   | 100          | 24    | A1718783 as58h08.x |
| 37         | 9     | 100.0   | 101          | 164   | BE152001 QV1-HT031 |
| 38         | 9     | 100.0   | 101          | 164   | BE179567 IL3-HT061 |
| 39         | 9     | 100.0   | 102          | 186   | BE365112 P11.24.B0 |
| 40         | 9     | 100.0   | 102          | 189   | T61718 yb92e07.r1  |
| 41         | 9     | 100.0   | 102          | 237   | AZ059959 RPCI-23-4 |
| 42         | 9     | 100.0   | 104          | 8     | AA518125 v123g11.r |
| 43         | 9     | 100.0   | 104          | 121   | AW810470 MR4-ST012 |
| 44         | 9     | 100.0   | 104          | 164   | BE145670 ILO-HT020 |
| 45         | 9     | 100.0   | 105          | 1     | AA014683 mh09g05.r |

Email: cgapbs-f@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/BTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM286 row: b column: 10.

## FEATURES

source  
1..40  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3617961"  
/clone\_lib="NIH\_MGC\_20"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: skin; Vector: pOTB7; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 12 a 8 c 11 g 9 t  
ORIGIN

Query Match 100.0% Score 9; DB 167; Length 40;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9  
|||||  
Db 11 CTAGTATTA 19

## RESULT 2

LOCUS A2769245 55 bp DNA GSS 16-FEB-2001  
DEFINITION IM0569010R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0569010 R, DNA sequence.  
ACCESSION A2769245  
VERSION A2769245.1 GI:12889180  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 55)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0569 row: 0 column: 10  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 55.  
Location/Qualifiers  
1..55  
/organism="Mus musculus"

## TITLE

JOURNAL  
COMMENT

## REFERENCE

AUTHORS  
TITLE  
JOURNAL  
COMMENT

## FEATURES

source  
1..55  
Location/Qualifiers  
/organism="Mus musculus"

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9  
|||||  
Db 50 CTAGTATTA 42

RESULT 4

AZ434302 57 bp DNA GSS 03-OCT-2000  
LOCUS 1M0220809R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0220809 R, DNA sequence.

ACCESSION AZ434302  
VERSION AZ434302.1 GI:10558315

KEYWORDS

SOURCE GSS.

ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 57)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

TITLE

Unpublished (2000)

JOURNAL

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00

Plate: 0220 row: E column: 09

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 57.

FEATURES

Location/Qualifiers

1..57

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0220809"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (gi14732114|gb|AF129072.1), a copy-number  
of pMD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 12 a 14 c 9 g 22 t  
ORIGIN

Query Match 100.0%; Score 9; DB 243; Length 57;

/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0569010"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 33 a 4 c 4 g 14 t  
ORIGIN

Query Match 100.0%; Score 9; DB 249; Length 55;  
Best Local Similarity 100.0%; Pred. NO. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTA 10

RESULT 3

AF097417/c 57 bp mRNA EST 10-NOV-1999  
LOCUS AF097417 Vespertilio superans library Vespertilio superans cDNA,  
mRNA sequence.

ACCESSION AF097417.1 GI:6341356

VERSION EST.

KEYWORDS Asian particolored bat.

SOURCE Vespertilio superans

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Chiroptera; Microchiroptera; Vespertilionidae;  
Vespertilio.

REFERENCE 1 (bases 1 to 57)

AUTHORS Kang,H.S. and Kim,J.W.

TITLE Differential Gene Expression of Active and Hibernating States in

JOURNAL Bats

COMMENT Unpublished (1999)

Contact: Kang HS

Anatomy Wonju College of Medicine, Yonsei University

Ilsandong 162, Wonju, Kangwon 220-701, Korea

Email: mhkim1@yml.yonsei.ac.kr.

Location/Qualifiers

1..57

/organism="Vespertilio superans"

/db\_xref="taxon:105273"

/clone\_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT 19 a 9 c 9 g 20 t

ORIGIN

Query Match 100.0%; Score 9; DB 14; Length 57;

Best Local Similarity 100.0%; Pred. NO. 3.2e+04;



Thu Oct 25 13:09:00 2001

**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
**JOURNAL** Tumor Gene Index  
**COMMENT** Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps@remail.nih.gov](mailto:cgaps@remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

MGI:1031143

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. .59

/organism="Mus musculus"

/strain="CZECH II"

/db\_xref="taxon:10090"

/clone="IMAGE:2650691"

/clone\_lib="NCI-CGAP\_Lu30"

/tissue\_type="tumor, metastatic to mammary"

/lab\_host="DHI08"

/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; transgenic model WNT-1, expression driven by

MMTV-LTR enhancer; Cloned unidirectionally. Primer: oligo

dT. Library constructed by Life Technologies.

Investigator providing samples: Gilbert Smith, NIH"

14 a ll c 7 g 27 t

**BASE COUNT****ORIGIN**

**Query Match** 100.0%; Score 9; DB 113; Length 59;  
**Best Local Similarity** 100.0%; Pred. No. 3.2e+04;  
**Matches** 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9

lllllllll

Db 5 CTAGTATTATA 13

**RESULT** 8

TA82F10Q/c

**LOCUS**

DEFINITION TA82F10Q 62 bp DNA GSS 13-DEC-2000

T. brucei sheared genomic DNA clone 82f10, reverse sequence,

genomic survey sequence.

ACCESSION AL459971

VERSION AL459971.1 GI:11860296

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 62)

Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,

Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,

Melville, S.E., Rajandream, M.A. and Barrell, B.G.

Direct Submission

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

Cambridge CB10 1SA, E-mail: [barrell@sanger.ac.uk](mailto:barrell@sanger.ac.uk) and[nhs@sanger.ac.uk](mailto:nhs@sanger.ac.uk)

Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared

to give a tight size distribution (

4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: [nelsayed@tigr.org](mailto:nelsayed@tigr.org)Details of T. brucei sequencing at the Sanger Centre are available  
 at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).**FEATURES**  
 source

Location/Qualifiers

1. .62

/organism="Trypanosoma brucei"

/strain="TREU927"

/db\_xref="taxon:5691"

/clone="82f10"

27 a 8 c 11 g 16 t

**BASE COUNT****ORIGIN**

Query Match 100.0%; Score 9; DB 258; Length 62;

Best Local Similarity 100.0%; Pred. No. 3.2e+04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 ctagtattata 9

lllllllll

Db 18 CTAGTATTATA 10

**RESULT** 9

AZ808515

**LOCUS**

DEFINITION AZ808515 66 bp DNA GSS 20-FEB-2001

2M0072A07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC2M0072A07 F, DNA sequence.

ACCESSION AZ808515

VERSION AZ808515.1 GI:12973933

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 66)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0072 row: A column: 07

Seq primer: CGTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 66.

Location/Qualifiers

1. .66

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0072A07"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi1473214|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed mouse DNA was annealed to chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 19 a 14 c 16 g 17 t  
ORIGIN

Query Match 100.0%; Score 9; DB 250; Length 66;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctactatata 9  
|||||||  
Db 34 CTACTATTA 42

RESULT 10  
AI761414/c  
LOCUS  
DEFINITION  
w965c06.x1 Soares NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
IMAGE:2369962 3' similar to TR:075425 075425 ORF3, SPLICEVARIANT\_B.  
// mRNA sequence.  
ACCESSION  
VERSION  
AI761414.1 GI:5177081  
SOURCE  
EST.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS  
1 (bases 1 to 67)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbbs@mail.nih.gov](mailto:cgapbbs@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Trace considered overall poor quality  
Insert Length: 1144 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1..67  
/organism="Homo sapiens"  
/strain="9606"  
/db\_xref="taxon:9606"  
/clones="IMAGE:2369962"  
/clone\_lib="Soares NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHP pool 1:  
309384-310919, 323208-325895 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HF8-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326663 Soares NBHOT  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 18 a 14 c 24 g 11 t

## ORIGIN

Query Match 100.0%; Score 9; DB 24; Length 67;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctactatata 9  
|||||||  
Db 18 CTACTATTA 10

RESULT 11  
AA250291/c  
LOCUS  
DEFINITION  
mz61402.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone  
IMAGE:717938 5', mRNA sequence.  
ACCESSION  
VERSION  
AA250291.1 GI:1882038  
KEYWORDS  
SOURCE  
EST.  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
AUTHORS  
1 (bases 1 to 71)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)

## JOURNAL

## COMMENT

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MG1:443434

Putative full length read

Vector to vector length is

Seq primer: -28ml3 rev2 ET from Amersham.

## FEATURES

## Source

1..71  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clones="IMAGE:717938"  
/clone\_lib="Soares mouse lymph node NBMLN"  
/sex="male"  
/tissue\_type="lymph node"  
/lab\_host="DH10B"  
/dev\_stage="4 weeks"  
/note="Organ: lymph node; Vector: p7T3D-Pac (Pharmacia)  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5].  
TGTTACCAATCTGAAGTGGAGCGCGGATACATTTTTTTTTTTTTTTTTTTT  
3'; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p7T3 vector. RNA  
provided by Dr. Bertrand Jordan. Library constructed and  
normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 32 a 10 c 12 g 17 t

## ORIGIN

Query Match 100.0%; Score 9; DB 4; Length 71;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu Oct 25 13:09:00 2001

```

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1081639
Seq primer: -40RP from Gibco.

FEATURES
Location/Qualifiers
1..71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3372035"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      32 a      10 c      12 g      17 t
ORIGIN

Query Match      100.0%; Score 9; DB 138; Length 71;
Best Local Similarity 100.0%; Pred. NO. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgtattata 9
|||||
Db 66 CTAGTATTA 58

RESULT 14
BE630111 71 bp mRNA EST 25-AUG-2000
LOCUS
IMAGE:3372035 3', mRNA sequence.
DEFINITION
BE630111
ACCESSION
BE630111
VERSION
BE630111.1 GI:9912799
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:939225
Seq primer: -40RP from Gibco.

FEATURES
Location/Qualifiers
1..71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1512373"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      32 a      10 c      12 g      17 t
ORIGIN

Query Match      100.0%; Score 9; DB 123; Length 71;
Best Local Similarity 100.0%; Pred. NO. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgtattata 9
|||||
Db 66 CTAGTATTA 58

RESULT 13
BE627067/c 71 bp mRNA EST 24-AUG-2000
LOCUS
IMAGE:3372035 5', mRNA sequence.
DEFINITION
BE627067
ACCESSION
BE627067.1 GI:9907487
VERSION
BE627067
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 17 a 12 c 11 g 31 t  
ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctagtattta 9  
|||||  
Db 5 CTAGTATTA 13

RESULT 15  
BE688187  
LOCUS  
DEFINITION BE688187 71 bp mRNA EST 11-SEP-2000  
IMAGE:3464723 3', mRNA sequence. Mus musculus cDNA clone  
ACCESSION BE688187  
VERSION  
KEYWORDS  
SOURCE EST.  
ORGANISM house mouse.  
Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 71)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:1384083.

FEATURES  
source  
1..71  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3464723"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 18 a 12 c 9 g 32 t  
ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;  
Best Local Similarity 100.0%; Pred. No. 3.2e+04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctagtattta 9  
|||||  
Db 5 CTAGTATTA 13

us-09-462-955-5.rst

---

Thu Oct 25 13:09:00 2001



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:28 ; Search time 393.52 Seconds  
(without alignments)  
36.699 Million cell updates/sec

Title: US-09-462-955-6

Perfect score: 23

Sequence: 1 ctgccagccagcagcctaggga 23

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSL1/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match % | Length | ID       | Description        |
|------------|-------|---------------|--------|----------|--------------------|
| 1          | 23    | 100.0         | 44     | AA02689  | CFDV-DNA fragment  |
| 2          | 23    | 100.0         | 44     | AA02687  | CFDV-DNA fragment  |
| C 3        | 19.8  | 86.1          | 324    | AA02685  | Human ORF5325      |
| C 4        | 19.8  | 86.1          | 1592   | AA027241 | Human secreted pro |
| C 5        | 18.8  | 81.7          | 524    | AA049239 | Human secreted pro |
| C 6        | 18.2  | 79.1          | 2512   | AA012867 | Human secreted pro |
| C 7        | 18.2  | 79.1          | 3131   | AA012867 | JM1-229 cell line  |
| C 8        | 17.8  | 77.4          | 496    | AA012867 | Gene 036 different |
| C 9        | 17.8  | 77.4          | 1658   | AA038123 | Human FATP alterna |
| C 10       | 17.8  | 77.4          | 2219   | AA018209 | Lung cancer associ |
| C 11       | 17.8  | 77.4          | 2222   | AA038122 | Human FATP protein |
|            |       |               |        | AA038124 | Human FATP variant |

|      |      |      |        |    |          |                    |
|------|------|------|--------|----|----------|--------------------|
| C 12 | 17.8 | 77.4 | 2816   | 19 | AA006641 | cDNA encoding huma |
| C 13 | 17.8 | 77.4 | 2818   | 21 | AA006641 | DNA encoding huma  |
| C 14 | 17.8 | 77.4 | 3217   | 20 | AA003365 | Nucleotide sequenc |
| C 15 | 17.8 | 77.4 | 3694   | 20 | AA003365 | Nucleotide sequenc |
| C 16 | 17.8 | 77.4 | 3704   | 20 | AA003365 | Nucleotide sequenc |
| C 17 | 17.8 | 77.4 | 6744   | 20 | AA038125 | Mutant lipase codi |
| C 18 | 17.2 | 74.8 | 1851   | 20 | AA027845 | Porcine BAC-P1Gf2- |
| C 19 | 17.2 | 74.8 | 5349   | 21 | AA065527 | Rhizobium species  |
| C 20 | 16.8 | 73.0 | 534720 | 19 | AA030458 | Rhizobium species  |
| C 21 | 16.8 | 73.0 | 536165 | 19 | AA030459 | Rhizobium species  |
| C 22 | 16.6 | 72.2 | 410    | 21 | AA002239 | Human secreted pro |
| C 23 | 16.6 | 72.2 | 446    | 21 | AA009796 | Human secreted pro |
| C 24 | 16.6 | 72.2 | 1662   | 20 | AA038080 | Human forkhead tra |
| C 25 | 16.6 | 72.2 | 1662   | 20 | AA038080 | Human forkhead tra |
| C 26 | 16.6 | 72.2 | 1845   | 20 | AA033613 | Human FKHL7 coding |
| C 27 | 16.6 | 72.2 | 1856   | 21 | AA076987 | Human breast tumou |
| C 28 | 16.6 | 72.2 | 2828   | 21 | AA014001 | DNA encoding huma  |
| C 29 | 16.6 | 72.2 | 3148   | 21 | AA076239 | Human ORF5 ORF2542 |
| C 30 | 16.6 | 72.2 | 3946   | 20 | AA038079 | Human forkhead tra |
| C 31 | 16.6 | 72.2 | 3946   | 20 | AA038079 | Human forkhead tra |
| C 32 | 16.4 | 71.3 | 300    | 20 | AA031671 | Human FKHL7 coding |
| C 33 | 16.4 | 71.3 | 846    | 20 | AA015019 | Human gene express |
| C 34 | 16.4 | 71.3 | 846    | 20 | AA015019 | DNA encoding trans |
| C 35 | 16.4 | 71.3 | 1210   | 20 | AA052236 | Protein PRO224 cDN |
| C 36 | 16.4 | 71.3 | 1210   | 21 | AA077549 | Human PRO224 cDN   |
| C 37 | 16.4 | 71.3 | 1210   | 21 | AA049721 | Human PRO224 cDN   |
| C 38 | 16.4 | 71.3 | 1210   | 22 | AA072394 | Human PRO224 cDN   |
| C 39 | 16.4 | 71.3 | 1264   | 21 | AA053830 | Human angiotensin  |
| C 40 | 16.4 | 71.3 | 1270   | 20 | AA011187 | Sequence encoding  |
| C 41 | 16.4 | 71.3 | 2206   | 15 | AA074445 | Gene encoding tran |
| C 42 | 16.4 | 71.3 | 2945   | 16 | AA087790 | DNA located 3' of  |
| C 43 | 16.4 | 71.3 | 3690   | 21 | AA072181 | S. lavenulae mcr   |
| C 44 | 16.4 | 71.3 | 3690   | 21 | AA072181 | Human low adenosin |
| C 45 | 16.4 | 71.3 | 4077   | 19 | AA035159 | Human adenosine re |
|      |      |      |        |    | AA028999 | Human endothelial  |

## ALIGNMENTS

RESULT 1  
ID AA02689 standard; DNA; 44 BP.  
XX AA02689;  
XX 10-MAY-1999 (first entry)  
XX CFDV-DNA fragment stem loop repeat region.  
DE Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;  
KW Yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.  
XX Coconut foliar decay virus  
XX Key Location/Qualifiers  
FH stem\_loop 4..22  
FT /\*tag= a  
FT /\*note= "RPT1"  
FT misc\_binding 4..10  
FT /\*tag= b  
FT /\*note= "Region binds to nucleotides 16 to 22"  
FT misc\_binding 16..22  
FT /\*tag= c  
FT /\*note= "Region binds to nucleotides 4 to 10"  
FT stem\_loop 27..43  
FT /\*tag= d  
FT /\*note= "RPT2"  
FT misc\_binding 27..32  
FT /\*tag= e  
FT /\*note= "Region binds to nucleotides 38 to 43"  
FT misc\_binding 36..43  
FT /\*tag= f  
FT /\*note= "Region binds to nucleotides 27 to 32"

```
XX PN DE19730502-A1.
XX XX
XX PD 21-JAN-1999.
XX PF 16-JUL-1997; 97DE-1030502.
XX PR 16-JUL-1997; 97DE-1030502.
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX DR WPI; 1999-096863/09.
XX PT Coconut foliar decay virus promoters - for gene expression in
XX PT bacteria and yeasts
XX PS Disclosure; Fig 4; 14pp; German.
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFDV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or
XX CC stable expression. Certain fragments of CFDV DNA have stronger promoter
XX CC activity in E. coli than the CaMV 35S promoter.
XX SQ Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;

Query Match 100.0%; Score 23; DB 20; Length 44;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaagcctggga 23
Db 1 ctgcccaggccgaagcctggga 23

RESULT 2
AAC02687
ID AAC02687 standard; DNA; 44 BP.
XX AC AAC02687;
XX DT 10-MAY-1999 (first entry)
XX DE CFDV-DNA fragment stem loop repeat region.
XX KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX OS Coconut foliar decay virus
XX PH Key Location/Qualifiers
FT FT stem_loop 4..22 a
FT FT /*tag= "RPT1"
FT FT misc_binding 4..10 b
FT FT /*tag= "RPT2"
FT FT misc_binding 16..22 c
FT FT /*tag= "Region binds to nucleotides 16 to 22"
FT FT /*tag= "Region binds to nucleotides 4 to 10"
FT FT stem_loop 27..43 d
FT FT /*tag= "RPT2"
FT FT misc_binding 27..32 e
FT FT /*tag= "Region binds to nucleotides 38 to 43"
FT FT misc_binding 38..43
```

```
FT FT /*tag= f
FT FT /note= "Region binds to nucleotides 27 to 32"
XX PN DE19730535-A1.
XX XX
XX PD 21-JAN-1999.
XX PF 16-JUL-1997; 97DE-1030535.
XX PR 16-JUL-1997; 97DE-1030535.
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX DR WPI; 1999-096867/09.
XX PT Coconut foliar decay virus promoters - for gene expression in
XX PT plants, bacteria and yeasts
XX PS Disclosure; Fig 4; 12pp; German.
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFDV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or
XX CC stable expression. Certain fragments of CFDV DNA have stronger promoter
XX CC activity in E. coli than the CaMV 35S promoter.
XX SQ Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;

Query Match 100.0%; Score 23; DB 20; Length 44;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaagcctggga 23
Db 1 ctgcccaggccgaagcctggga 23

RESULT 3
AAC75325/c
ID AAC75325 standard; cDNA; 324 BP.
XX AC AAC75325;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF880 polynucleotide sequence SEQ ID NO:1759.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
XX KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX KW neurodegenerative disorder; osteoarthritis; hypothyroidism; SCID; AIDS;
XX KW cardiovascular disease; diabetes mellitus; erythematous; infection;
XX KW cholesterol ester storage; systemic lupus erythematosus; asthma;
XX KW severe combined immunodeficiency; malaria; autoimmune disorder;
XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX KW thrombosis; cartilage damage; antinflammatory disease; coagulation;
XX OS Homo sapiens.
XX PN WO200058473-A2.
XX PR 05-OCT-2000.
```

```

XX PF 31-MAR-2000; 2000WO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX XX Shimkets RA, Leach M;
XX PI WPI: 2000-602362/57.
XX XX P-PSDB; AAB41116.
XX XX Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS Claim 5; Page 1386-1387; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antipariatic; antiparkinsonian; nootropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX CC antithyroid; and antianemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy.
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, cholesterol ester storage, systemic lupus
XX CC erythematosus, malaria, autoimmune disorders, asthma,
XX CC allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX XX Sequence 324 BP; 82 A; 84 C; 89 G; 69 T; 0 other;
XX SQ

Query Match 86.1%; Score 19.8; DB 21; Length 324;
Best Local Similarity 91.3%; Pred. No. 46;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaaggcctggga 23
   |||||
DB 71 CTGCCAGGCCGAGTGCTGGGA 49

RESULT 4
AAZ27241/c
ID AAZ27241 standard; cDNA; 1592 BP.
XX AAZ27241;
AC AAZ27241;
XX 23-NOV-1999 (first entry)
XX Human secreted protein cDNA encoding gene 9.
XX Secreted protein; human; gene therapy; diagnosis; treatment; cancer;
XX protein therapy; tumor; neurodegenerative disorder; blood disorder; AIDS;
XX developmental abnormality; leukemia; immune system; autoimmune disease;
XX hepatic disease; renal disease; inflammation; allergy; schizophrenia;
XX Alzheimer's disease; cognitive disorder; arthritis; infection; psoriasis;
XX transplant rejection; diabetes; asthma; sepsis; acne; metabolic disorder;
XX cardiovascular disorder; food additive; preservative; ss.
XX XX

Query Match 86.1%; Score 19.8; DB 21; Length 324;
Best Local Similarity 91.3%; Pred. No. 46;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaaggcctggga 23
   |||||
DB 71 CTGCCAGGCCGAGTGCTGGGA 49

RESULT 4
AAZ27241/c
ID AAZ27241 standard; cDNA; 1592 BP.
XX AAZ27241;
AC AAZ27241;
XX 23-NOV-1999 (first entry)
XX Human secreted protein cDNA encoding gene 9.
XX Secreted protein; human; gene therapy; diagnosis; treatment; cancer;
XX protein therapy; tumor; neurodegenerative disorder; blood disorder; AIDS;
XX developmental abnormality; leukemia; immune system; autoimmune disease;
XX hepatic disease; renal disease; inflammation; allergy; schizophrenia;
XX Alzheimer's disease; cognitive disorder; arthritis; infection; psoriasis;
XX transplant rejection; diabetes; asthma; sepsis; acne; metabolic disorder;
XX cardiovascular disorder; food additive; preservative; ss.
XX XX

OS Homo sapiens.
XX WO9946289-A1.
XX PD 16-SEP-1999.
XX PF 11-MAR-1999; 99WO-US05721.
XX PR 12-MAR-1998; 98US-0077686.
XX PR 12-MAR-1998; 98US-0077687.
XX PR 12-MAR-1998; 98US-0077696.
XX PR 12-MAR-1998; 98US-0077714.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Ferrie AM, Rosen CA, Florence C, Young PE, Yu G, Ni J;
XX WPI: 1999-551363/46.
XX DR P-PSDB; AAY45265.
XX DR New isolated human genes, useful for diagnosis and treatment of, e.g.
XX PT cancers -
XX PS Claim 1a; 155; 306pp; English.
XX CC This invention describes novel human genes and the secreted proteins
XX CC they encode. The polynucleotides and their corresponding secreted
XX CC polypeptides are useful for preventing, treating or ameliorating medical
XX CC conditions e.g. by protein or gene therapy. Also pathological conditions
XX CC can be diagnosed by determining the amount of the new polypeptides in a
XX CC sample or by determining the presence of mutations in the new
XX CC polynucleotides. Specific uses are described for each of the
XX CC polynucleotides of the invention, based on which tissues they are most
XX CC highly expressed in, and include developing products for the diagnosis or
XX CC treatment of cancer, tumors, neurodegenerative disorders, developmental
XX CC abnormalities, blood disorders, leukemias, diseases of the immune system,
XX CC autoimmune diseases, hepatic and renal disease, inflammation, allergies,
XX CC Alzheimer's and cognitive disorders, schizophrenia, arthritis,
XX CC infections, AIDS, transplant rejection, diabetes, asthma, sepsis, acne,
XX CC psoriasis, cardiovascular disorders, and metabolic disorders. The
XX CC polypeptides or polynucleotides can also be used as food additives or
XX CC preservatives. The polypeptides are also useful for identifying their
XX CC binding partners. AAZ27233-227265 encode human secreted proteins
XX CC described in the method of the invention.
XX SQ Sequence 1592 BP; 432 A; 382 C; 418 G; 360 T; 0 other;

Query Match 86.1%; Score 19.8; DB 20; Length 1592;
Best Local Similarity 91.3%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaaggcctggga 23
   |||||
DB 1066 CTGCCAGGCCGAGGCTGAGA 1044

RESULT 5
AAZ49239/c
ID AAZ49239 standard; cDNA; 524 BP.
XX AAZ49239;
AC AAZ49239;
XX 07-MAR-2000 (first entry)
XX Human hydrolase homologue HHH-6 cDNA fragment 1514414F6.
XX Hydrolase; homologue; HHH-6; N-terminal asparagine amidohydrolase;
XX vanin-1; glycosyl hydrolase; glucosylase;
XX N-acetylglucosamine 6-P deacetylase; diagnosis;
XX treatment; prevention; expression; disorder; carbohydrate; metabolism;
XX antagonist; reproductive disorder; cell proliferation; ds.
XX XX

```

[illegible]

```

FT CDS 449...2668
EX /*tag= a
PN W09733551-A2.
XX
PD 18-SEP-1997.
XX
PF 14-MAR-1997; 97WO-US04191.
XX
PR 15-MAR-1996; 96US-0013438.
XX
XX (MILL-) MILLENNIUM PHARM.
PA
PI Shyjan AW;
XX
DR WPI; 1997-470615/43.
DR P-PSDB; AAW25790.
XX
XX Gene 036 with reduced or amplified expression in tumour cells -
PT used, optionally with genes 097, 030 and 056 or their protein
PT products, for diagnosis and treatment of colonic cancer
XX
XX Claim 19; Fig 2; 117pp; English.
XX
CC Gene 036, the coding sequence for which is claimed, was identified
CC by differential display analysis as being expressed at a higher
CC level in normal colon tissues than in cancerous colon tissues.
CC Gene 036 is a candidate tumour suppressor gene. It shows 96%
CC sequence identity to EST sequence clone B4507, and has been
CC assigned to human chromosome 4. It encodes a 739-amino acid
CC protein (see AAW24790). A correlation was found between an increase
CC in the expression level of gene 036 and a decrease in a colon
CC cell's tumour potential. Hence, methods that increase the level
CC of expression of gene 036 may inhibit or slow the progression
CC to tumours and cancer, e.g. colon cancer. Genes 030 (see AAT91703)
CC and 056 (see AAT91705) show similar expression patterns to 036, in
CC colon tumour tissues than in normal colon tissues. The identified
CC genes can be used diagnostically or as targets for therapy
CC (including gene therapy), or can be used to identify compounds
CC useful in the diagnosis, prevention and treatment of tumours and
XX cancers.
XX
SQ Sequence 3131 BP; 876 A; 831 C; 746 G; 676 T; 2 other;

Query Match 79.1%; Score 18.2; DB 18; Length 3131;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ctgccaggccgaaggcctggga 23
DB 957 ctgccagaccgagtgctgaga 979

RESULT 8
AAZ38123/C
ID AAZ38123 standard; cDNA; 496 BP.
XX
AC AAZ38123;
XX
DT 08-FEB-2000 (first entry)
XX
DE Human FATP alternatively spliced sequence.
XX
KW Fatty acid transport protein; FATP; hfATP; cardiomyopathy; diabetes;
KW long-chain fatty acid metabolism; obesity; human; ss.
XX
OS Homo sapiens.
XX
PN W09951740-A2.
XX
PD 14-OCT-1999.

```

```

XX 02-APR-1999; 99WO-EP02295.
PF
XX 06-APR-1998; 98EP-0400823.
PR
XX (JANC ) JANSSEN PHARM NV.
XX (UNIW ) UNIV WASHINGTON.
PA
XX Martin G, Nemoto M, Desb SS, Auwerx J;
PI
XX WPI; 1999-620202/53.
DR
XX New human fatty acid transport protein, hfATP, useful to screen for
XX inhibitors or enhancers useful to regulate fatty acid metabolism -
PT
XX Claim 2; Fig 1A-D; 83pp; English.
PS
XX The invention provides a human fatty acid transport protein (hfATP).
XX hfATP is believed to be involved in the modulation long-chain fatty acid
XX metabolism; the protein and polynucleotides therefore enable production
XX of compositions comprising a component regulating (inhibiting or
XX enhancing) expression of the hfATP gene useful therapeutically to alter
XX intracellular or blood levels of long chain fatty acids. Such compounds
XX are especially useful to treat conditions associated with deficient
XX regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or
XX diabetes or an enhancer to treat obesity. The polynucleotides are also
XX useful to screen compounds for their effects on hfATP expression, e.g.
XX by measuring mRNA transcription in cells/cell extracts (e.g. liver
XX cells) contacted with the compound and comparing with that in non-
XX contacted cells. Sequences AAZ38122-123 represent nucleotide sequence of
XX clones containing 2 kb sequences obtained from a cDNA library of human
XX adipose tissue. The present sequence represents an alternative spliced
XX sequence found in the analyzed clones.
XX
SQ Sequence 496 BP; 56 A; 151 C; 202 G; 87 T; 0 other;

Query Match 77.4%; Score 17.8; DB 20; Length 496;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
DB 435 TGCCCCAGGCAGAGGCCAGGG 415

RESULT 9
AAF18209/C
ID AAF18209 standard; DNA; 1658 BP.
XX
AC AAF18209;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 228.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN W0200055180-A2.
XX
XX 21-SEP-2000.
PD
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA

```

Thu Oct 25 13:09:03 2001

us-09-462-955-6.rng

PA (ROSE/) ROSEN C A.  
 XX Ruben SM;  
 PI WPI; 2000-587514/55.  
 DR P-PSDB; AAB58333.  
 XX Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer -  
 XX  
 PS Claim 1; Page 689-690; 1425pp; English.  
 XX Polynucleotide sequences AAF17962 - AAF18424 encode human lung cancer  
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
 CC associated proteins and polynucleotide sequences, their agonists, and  
 CC antagonists may have neuroprotective; cytostatic; cardioactive;  
 CC immunomodulatory; muscular active general; gynecological; or antibacterial  
 CC general; nephrotropic; anti-infective; cytostatic; cardioactive;  
 CC activity. The invention also includes antibodies specific for the  
 CC protein or polynucleotide sequences. The lung cancer associated  
 CC polynucleotide sequences may be used for detection of lung cancer,  
 CC chromosome identification, as chromosome markers, and for numerous other  
 CC diagnostic or research purposes. The proteins may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders. The proteins may also be used in the treatment of wounds and  
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
 CC peptide AAB5849 are used in the course of the invention for the  
 CC identification and characterization of the polynucleotide and protein  
 CC sequences.  
 XX  
 SQ Sequence 1658 BP; 386 A; 429 C; 374 G; 463 T; 6 other;

Query Match 77.4%; Score 17.8; DB 21; Length 1658;  
 Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ctgccagggccgaagcctgg 21  
 ||||| ||||| ||||| |||||  
 DB 674 CTGCCAGGTAGAGGCGCTGG 654

RESULT 10  
 AAZ38122/c  
 ID AAZ38122 standard; cDNA; 2219 BP.  
 XX AC AAZ38122;  
 XX 08-FEB-2000 (first entry)  
 DT Human FATP protein encoding cDNA.  
 DE Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;  
 KW long-chain fatty acid metabolism; obesity; human; ss.  
 KW Homo sapiens.  
 XX WO9951740-A2.  
 XX 14-OCT-1999.  
 XX 02-APR-1999; 99WO-EP02295.  
 XX 06-APR-1998; 98EP-0400823.  
 XX (JANC ) JANSSEN PHARM NV.  
 XX (UNIW ) UNIV WASHINGTON.  
 XX Martin G, Nemoto M, Deeb SS, Auwerx J;  
 WPI; 1999-620202/53.

P-PSDB; AAY40435.  
 New human fatty acid transport protein, hFATP, useful to screen for  
 inhibitors or enhancers useful to regulate fatty acid metabolism -  
 Claim 2; Fig 1A-D; 83pp; English.  
 The invention provides a human fatty acid transport protein (hFATP).  
 hFATP is believed to be involved in the modulation long-chain fatty acid  
 metabolism; the protein and polynucleotides therefore enable production  
 of compositions comprising a component regulating (inhibiting or  
 enhancing) expression of the hFATP gene useful therapeutically to alter  
 intracellular or blood levels of long chain fatty acids. Such compounds  
 are especially useful to treat conditions associated with deficient  
 regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or  
 diabetes or an enhancer to treat obesity. The polynucleotides are also  
 useful to screen compounds for their effects on hFATP expression, e.g.  
 by measuring mRNA transcription in cells/cell extracts (e.g. liver  
 cells) contacted with the compound and comparing with that in non-  
 contacted cells. Sequences AAZ38122-123 represent nucleotide sequence of  
 clones containing 2 kb sequences obtained from a cDNA library of human  
 adipose tissue. The present sequence represents a cDNA encoding the  
 hFATP protein.  
 Sequence 2219 BP; 369 A; 710 C; 731 G; 406 T; 3 other;

Query Match 77.4%; Score 17.8; DB 20; Length 2219;  
 Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tccccagggccgaagcctgg 22  
 ||||| ||||| ||||| |||||  
 DB 539 TCCCAGGCGAGGCGCCAGG 519

RESULT 11  
 AAZ38124/c  
 ID AAZ38124 standard; cDNA; 2222 BP.  
 XX AC AAZ38124;  
 XX 08-FEB-2000 (first entry)  
 DT Human FATP variant nucleotide sequence.  
 DE Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;  
 KW long-chain fatty acid metabolism; obesity; human; variant; ss.  
 KW Homo sapiens.  
 XX WO9951740-A2.  
 XX 14-OCT-1999.  
 XX 02-APR-1999; 99WO-EP02295.  
 XX 06-APR-1998; 98EP-0400823.  
 XX (JANC ) JANSSEN PHARM NV.  
 XX (UNIW ) UNIV WASHINGTON.  
 XX Martin G, Nemoto M, Deeb SS, Auwerx J;  
 WPI; 1999-620202/53.

New human fatty acid transport protein, hFATP, useful to screen for  
 inhibitors or enhancers useful to regulate fatty acid metabolism -  
 Claim 2; Fig 3; 83pp; English.  
 The invention provides a human fatty acid transport protein (hFATP).  
 hFATP is believed to be involved in the modulation long-chain fatty acid

CC metabolism; the protein and polynucleotides therefore enable production  
 CC of compositions comprising a component regulating (inhibiting or  
 CC enhancing) expression of the hFATP gene useful therapeutically to alter  
 CC intracellular or blood levels of long chain fatty acids. Such compounds  
 CC are especially useful to treat conditions associated with deficient  
 CC regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or  
 CC diabetes or an enhancer to treat obesity. The polynucleotides are also  
 CC useful to screen compounds for their effects on hFATP expression, e.g.  
 CC by measuring mRNA transcription in cells/cell extracts (e.g. liver  
 CC cells) contacted with the compound and comparing with that in non-  
 CC contacted cells. The present sequence represents a variant of the  
 CC nucleotide sequence encoding the hFATP protein, obtained from a cDNA  
 CC library of human adipose tissue.  
 XX  
 SQ Sequence 2222 BP; 393 A; 666 C; 696 G; 464 T; 3 other;

Query Match 77.4%; Score 17.8; DB 20; Length 2222;  
 Best Local Similarity 90.5%; Pred. No. 2.2e-02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgccagggccgaaggcctgg 22  
 ||||| ||||| ||||| |||||  
 Db 539 TGCCAGGCAGAGCCAGGG 519

## RESULT 12

AAV00641/c  
 ID AAV00641 standard; cDNA; 2816 BP.

XX AC AAV00641;

XX DT 27-MAR-1998 (first entry)

XX DE cDNA encoding human endothelial PAS domain protein 1 (EPAS1).

XX KW Endothelial PAS domain protein 1; EPAS1; screening assay; gene therapy;  
 KW endothelial tissue specification; EPAS1 binding; heat shock protein;  
 KW basic helix-loop-helix motif; bHLH; hypoxia inducible factor;  
 KW HIF-1 alpha binding site; binding affinity; genetic hybridisation screen;  
 KW immunogen; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX CDS 150...2760  
 XX FT /\*tag= a

XX PN US5695963-A.

XX PD 09-DEC-1997.

XX PF 17-JAN-1997; 97US-0785241.

XX PR 17-JAN-1997; 97US-0785241.

XX PA (TEXA ) UNIV TEXAS SYSTEM.

XX PI McKnight SL, Russell DW, Tian H;

XX DR WPI: 1998-041300/04.

XX DR P-PSDB; AAW37097.

XX PT Endothelial PAS domain protein 1 proteins - used for isolating EPAS1  
 XX regulators, e.g. a heat shock protein

XX PS Claim 9; Columns 19-22; 23pp; English.

XX CC The present sequence encodes a novel human endothelial PAS domain  
 CC protein 1 (EPAS1). EPAS1 proteins can regulate specification of  
 CC endothelial tissue, such as vasculature, the blood-brain barrier, etc..  
 CC The protein can be used in a screening assay for agents that  
 CC modulate binding of EPAS1 to a binding target, especially a basic

CC helix-loop-helix (bHLH)/PAS protein, a heat shock protein or hypoxia  
 CC inducible factor (HIF-1 alpha) binding site. The EPAS1 protein is  
 CC incubated with the binding target and a test agent and the effect of the  
 CC test agent on the binding affinity of the protein for the target is  
 CC determined. The proteins may be produced recombinantly from transformed  
 CC host cells from the subject EPAS1 encoding nucleic acids or purified from  
 CC mammalian cells. The proteins may be used in diagnosis (e.g. genetic  
 CC hybridisation screens for EPAS1 transcripts), therapy (e.g. gene therapy  
 CC to modulate EPAS1 gene expression) and in the biopharmaceutical  
 CC industry (e.g. as immunogens, reagents for isolating B-cell specific  
 CC activators or other transcriptional regulators, reagents for screening  
 CC chemical libraries for lead pharmacological agents, etc.).  
 XX

SQ Sequence 2816 BP; 669 A; 880 C; 764 G; 503 T; 0 other;

Query Match 77.4%; Score 17.8; DB 19; Length 2816;  
 Best Local Similarity 90.5%; Pred. No. 2.1e-02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagggccgaaggcctgg 21  
 ||||| ||||| ||||| |||||

Db 2784 CTGCCAGGTAGAGGCTGG 2764

## RESULT 13

AAZ94053/c

ID AAZ94053 standard; cDNA; 2818 BP.

XX AC AAZ94053;

XX DT 05-JUN-2000 (first entry)

XX DE DNA encoding human endothelial PAS domain protein-1.

XX KW Endothelial PAS domain protein-1; EPAS1; human; angiogenesis;  
 KW antiarteriosclerotic; antitumour; atherosclerosis; tumour;  
 KW vascular disease; vulnary; cardiant; vasotropic;  
 KW cerebroprotective; gene therapy; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX CDS 150...2762  
 XX FT /\*tag= a

XX PN WO200009657-A2.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18539.

XX PR 14-AUG-1998; 98US-0096515.

XX PA (HARD ) HARVARD COLLEGE.

XX PI Lee M, Maemura K, Hiesh C;

XX DR WPI: 2000-205996/18.

XX DR P-PSDB; AAY79161.

XX PT Modulation of angiogenesis in mammals, useful for treating e.g.  
 XX atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue  
 XX damage, ischemia, balloon angioplasty, frostbite, gangrene or poor  
 XX circulation

XX PS Disclosure; Page 14-15; 57pp; English.

XX CC The present sequence is that of cDNA encoding human endothelial PAS  
 CC domain protein-1 (EPAS1, see AAY79161). The invention is based on  
 CC the discovery that EPAS1 binds to cis-acting regulatory sequences  
 CC associated with genes encoding angiogenic factors such as vascular  
 CC endothelial cell growth factor (VEGF) and VEGF receptors such as

Thu Oct 25 13:09:03 2001

us-09-462-955-6.rng

CC KDR/flk-1 and flt-1, thereby transactivating the promoters of such  
 CC genes. A claimed method of inhibiting angiogenesis in a mammal  
 CC comprises administering to the mammal a compound which inhibits  
 CC binding of EPAS1 to the cis-acting transcription regulatory DNA of  
 CC an angiogenic factor (see AA294051). The compound may be an EPAS1  
 CC polypeptide lacking a transactivation domain (see AA79160) or a  
 CC nucleic acid encoding such a polypeptide, or an antisense nucleic acid  
 CC complementary to mRNA encoding EPAS1, or an EPAS1-specific antibody.  
 CC The compound is preferably administered to an atherosclerotic  
 CC lesion or to a tumour site. Angiogenesis is also inhibited using a  
 CC compound, such as an EPAS1 dominant negative mutant, which inhibits  
 CC binding of EPAS1 to the EPAS-binding element, ABNT4 (see AA79162).  
 CC Angiogenesis can be promoted by administering EPAS1 DNA to increase  
 CC expression of VEGF or VEGF receptor in endothelial cells of a  
 CC patient suffering from peripheral vascular disease, cerebral  
 CC vascular disease, hypoxic tissue damage (e.g. hypoxic damage to  
 CC heart tissue), or coronary vascular disease as well as to treat  
 CC patients who have, or have had, transient ischaemic attacks,  
 CC vascular graft surgery, balloon angioplasty, frostbite, gangrene,  
 CC or poor circulation.  
 CC  
 CC Sequence 2818 BP; 669 A; 882 C; 764 G; 503 T; 0 other;

Query Match 77.4%; Score 17.8; DB 21; Length 2818;  
 Best Local Similarity 90.5%; Pred. No. 2.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagcgaagcctgg 21  
 |||||  
 Db 2784 CTGCCAGGTAGAGGCTGG 2764

RESULT 14  
 AA200365/C  
 ID AA200365 standard; DNA; 3217 BP.  
 XX  
 AC AA200365;  
 DT 26-OCT-1999 (first entry)  
 DE Nucleotide sequence of murine mmFATP1.  
 XX  
 XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;  
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9936537-A2.  
 XX  
 PD 22-JUL-1999.

XX 14-JAN-1999; 99WO-US00182.  
 XX  
 XX 14-JAN-1999; 99US-0232201.  
 PR 15-JAN-1998; 98US-0071374.  
 PR 20-JUL-1998; 98US-0093491.  
 PR 04-DEC-1998; 98US-0110941.  
 PR 14-JAN-1999; 99US-0232195.  
 PR 14-JAN-1999; 99US-0232197.  
 PR 14-JAN-1999; 99US-0232200.

XX (MILL-) MILLENNIUM PHARM INC.  
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.  
 XX  
 XX Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;  
 XX  
 XX WPI; 1999-444398/37.  
 DR P-PSDB; AAY14955.  
 XX

XX Fatty acid transport proteins and related polynucleotides, useful  
 PT for treating obesity, diabetes and heart disease  
 XX  
 XX Claim 31; Fig 44A-C; 255pp; English.

PS Example 1; Fig 62; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)  
 CC that mediate transport of long chain fatty acids (LCFAs) across cell  
 CC membranes into cells. Human and murine FATP proteins and nucleic acids  
 CC encoding the proteins are provided. The FATP proteins can be produced  
 CC by standard recombinant methodology. Fatty acid uptake by cells can be  
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.  
 CC In particular, antisense oligonucleotides can be used to modulate FATP  
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid  
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac  
 CC muscle or liver by administration of a complex of the agent and a FATP6  
 CC binding moiety. DNA encoding FATP proteins can be used as a reference  
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake  
 CC by administering an inhibitor or enhancer of FATP transport function in  
 CC the small intestine can decrease or increase calories available as fats,  
 CC and can decrease or increase circulating fatty acids. Blocking the  
 CC function of FATP4 and also FATP2, is useful for treating obesity,  
 CC diabetes and heart disease.  
 XX  
 SQ Sequence 3217 BP; 616 A; 932 C; 911 G; 758 T; 0 other;

Query Match 77.4%; Score 17.8; DB 20; Length 3217;  
 Best Local Similarity 90.5%; Pred. No. 2.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcacagcgaagcctgg 22  
 |||||  
 Db 517 TGCCAGGCAGAGCCAGG 497

RESULT 15  
 AA200356/C  
 ID AA200356 standard; DNA; 3694 BP.  
 XX  
 AC AA200356;  
 DT 26-OCT-1999 (first entry)  
 DE Nucleotide sequence of human hsfATP1.  
 XX  
 XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;  
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9936537-A2.  
 XX  
 PD 22-JUL-1999.

XX 14-JAN-1999; 99WO-US00182.  
 XX  
 XX 14-JAN-1999; 99US-0232201.  
 PR 15-JAN-1998; 98US-0071374.  
 PR 20-JUL-1998; 98US-0093491.  
 PR 04-DEC-1998; 98US-0110941.  
 PR 14-JAN-1999; 99US-0232195.  
 PR 14-JAN-1999; 99US-0232197.  
 PR 14-JAN-1999; 99US-0232200.

XX (MILL-) MILLENNIUM PHARM INC.  
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.  
 XX  
 XX Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;  
 XX  
 XX WPI; 1999-444398/37.  
 DR P-PSDB; AAY14946.  
 XX  
 XX Fatty acid transport proteins and related polynucleotides, useful  
 PT for treating obesity, diabetes and heart disease  
 XX  
 XX Claim 31; Fig 44A-C; 255pp; English.



XX

CC The invention provides a family of fatty acid transport proteins (FATPs)  
 CC that mediate transport of long chain fatty acids (LCFAs) across cell  
 CC membranes into cells. Human and murine FATP proteins and nucleic acids  
 CC encoding the proteins are provided. The FATP proteins can be produced  
 CC by standard recombinant methodology. Fatty acid uptake by cells can be  
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.  
 CC In particular, antisense oligonucleotides can be used to modulate FATP  
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid  
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac  
 CC muscle or liver by administration of a complex of the agent and a FATP6  
 CC binding moiety. DNA encoding FATP proteins can be used as a reference  
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake  
 CC by administering an inhibitor or enhancer of FATP transport function in  
 CC the small intestine can decrease or increase calories available as fats,  
 CC and can decrease or increase circulating fatty acids. Blocking the  
 CC function of FATP4 and also FATP2, is useful for treating obesity,  
 CC diabetes and heart disease.

XX

SQ Sequence 3694 BP; 622 A; 1202 C; 1110 G; 760 T; 0 other;

Query Match

Best Local Similarity 77.4%; Score 17.8; DB 20; Length 3694;

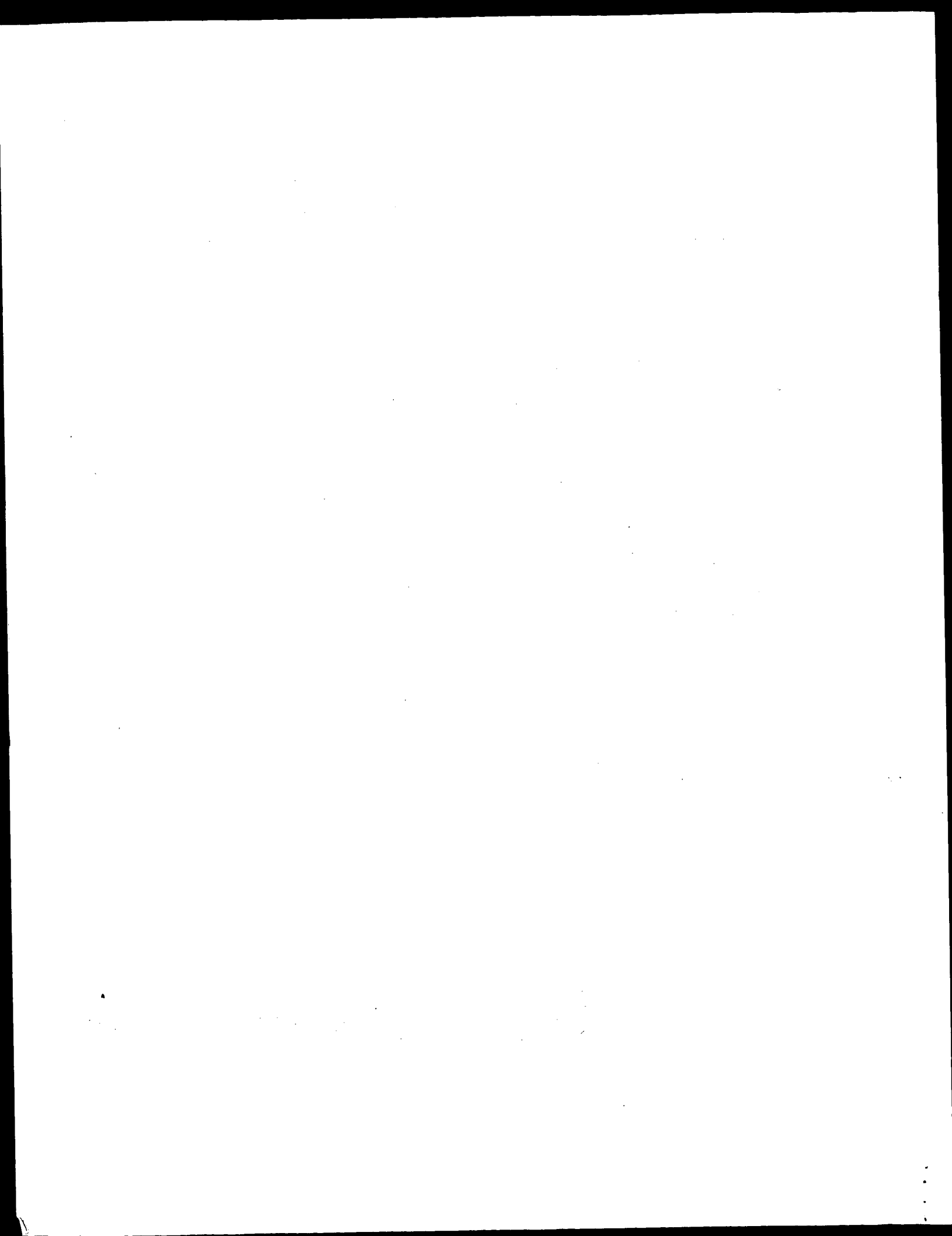
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccagggccgaagcctggg 22

||||||| ||||| |||

Db 691 TGCCACAGCAGAGGCGAGGG 671

Search completed: October 24, 2001, 10:07:30  
 Job time: 635 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:59 ; Search time 5479.82 Seconds  
(without alignments)  
39.676 Million cell updates/sec

Title: US-09-462-955-6

Perfect score: 23

Sequence: 1 ctgccagccgcaagcctggga 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_estl1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query % | Match | Length | DB       | ID        | Description |
|------------|-------|---------|-------|--------|----------|-----------|-------------|
| C 1        | 19.8  | 86.1    | 935   | 152    | BG327577 | BG327577  | 602426566   |
| C 2        | 19.8  | 86.1    | 959   | 144    | BF119646 | BF119646  | 601757514   |
| C 3        | 19.4  | 84.3    | 405   | 117    | AW522130 | UI-R-BJOP | AW522130    |
| C 4        | 19.4  | 84.3    | 512   | 19     | AI406571 | AI406571  | EST234857   |
| C 5        | 19.4  | 84.3    | 517   | 148    | BF403937 | BF403937  | UI-R-CAI-   |
| C 6        | 18.8  | 81.7    | 176   | 141    | BE860673 | BE860673  | UI-M-AH1-   |
| C 7        | 18.8  | 81.7    | 238   | 126    | BE147701 | BE147701  | BB147701    |
| C 8        | 18.8  | 81.7    | 313   | 114    | AW336731 | AW336731  | 21454 MAR   |
| C 9        | 18.8  | 81.7    | 393   | 130    | BE284196 | BE284196  | BB284196    |
| C 10       | 18.8  | 81.7    | 462   | 169    | BF775369 | BF775369  | 285513 MA   |
| C 11       | 18.4  | 80.0    | 438   | 3      | AA158436 | AA158436  | zo59h11.r   |
| C 12       | 18.4  | 80.0    | 455   | 117    | AW553523 | AW553523  | L0228C11-   |
| C 13       | 18.2  | 79.1    | 286   | 161    | BB573895 | BB573895  | BB573895    |
| C 14       | 18.2  | 79.1    | 317   | 139    | BE757761 | BE757761  | 212291 MA   |
| C 15       | 18.2  | 79.1    | 318   | 241    | AZ287293 | RPCI-23-1 | AZ287293    |
| C 16       | 18.2  | 79.1    | 377   | 147    | BF370922 | BF370922  | MR0-FN009   |
| C 17       | 18.2  | 79.1    | 392   | 21     | A1520113 | A1520113  | LD40089.5   |
| C 18       | 18.2  | 79.1    | 453   | 172    | BF993725 | BF993725  | IL5-GN017   |
| C 19       | 18.2  | 79.1    | 461   | 115    | AW366574 | AW366574  | RC2-HT014   |
| C 20       | 18.2  | 79.1    | 464   | 104    | AI949053 | AI949053  | wg36f03.x   |
| C 21       | 18.2  | 79.1    | 512   | 22     | AI565947 | AI565947  | tg68e06.x   |
| C 22       | 18.2  | 79.1    | 542   | 121    | AW80881  | QV0-OT003 | QV0-OT003   |
| C 23       | 18.2  | 79.1    | 556   | 229    | AQ479885 | AQ479885  | RPCI-11-2   |
| C 24       | 18.2  | 79.1    | 773   | 168    | BF700908 | BF700908  | 602128051   |
| C 25       | 18.2  | 79.1    | 886   | 151    | BF666792 | BF666792  | 602121104   |
| C 26       | 18.2  | 79.1    | 1067  | 221    | CNS03NKE | CNS03NKE  | Tetraodon   |
| C 27       | 17.8  | 77.4    | 243   | 107    | AU058457 | AU058457  | AU058457    |
| C 28       | 17.8  | 77.4    | 295   | 121    | AW877400 | AW877400  | MR4-PT005   |
| C 29       | 17.8  | 77.4    | 345   | 168    | BF721502 | BF721502  | mab17h10    |
| C 30       | 17.8  | 77.4    | 347   | 20     | AI447502 | AI447502  | mp68c06.x   |
| C 31       | 17.8  | 77.4    | 349   | 167    | BE385251 | BE385251  | 601277687   |
| C 32       | 17.8  | 77.4    | 363   | 168    | BF722349 | BF722349  | mab17h10    |
| C 33       | 17.8  | 77.4    | 370   | 140    | BE771160 | BE771160  | RC5-F007    |
| C 34       | 17.8  | 77.4    | 371   | 147    | BF372847 | BF372847  | RC5-F007    |
| C 35       | 17.8  | 77.4    | 372   | 140    | BE771169 | BE771169  | RC5-F007    |
| C 36       | 17.8  | 77.4    | 384   | 115    | AW416678 | AW416678  | 52227 MAR   |
| C 37       | 17.8  | 77.4    | 390   | 144    | BF087382 | BF087382  | QV2-HT054   |
| C 38       | 17.8  | 77.4    | 397   | 18     | AI326029 | AI326029  | mr30e09.x   |
| C 39       | 17.8  | 77.4    | 397   | 115    | AW408256 | AW408256  | UI-HF-BK0   |
| C 40       | 17.8  | 77.4    | 401   | 121    | AW38277  | AW38277   | QV2-LT005   |
| C 41       | 17.8  | 77.4    | 409   | 141    | BE997489 | BE997489  | 601432319   |
| C 42       | 17.8  | 77.4    | 412   | 19     | AI386449 | AI386449  | mr30e09.y   |
| C 43       | 17.8  | 77.4    | 418   | 140    | BE818183 | BE818183  | RC6-BN027   |
| C 44       | 17.8  | 77.4    | 430   | 162    | BE014102 | BE014102  | 125686 MA   |
| C 45       | 17.8  | 77.4    | 433   | 162    | BE013607 | BE013607  | 125013 MA   |

## ALIGNMENTS

RESULT 1  
 BG327577/c  
 LOCUS BG327577 935 bp mRNA EST 27-FEB-2001  
 DEFINITION 602426566F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4564192 5', mRNA sequence.  
 ACCESSION BG327577  
 VERSION BG327577.1 GI:13134015  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLC1279 row: d column: 17  
 High quality sequence stop: 789.

## FEATURES

source  
 Location/Qualifiers  
 1..935  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4564192"  
 /clone\_lib="NIH\_MGC\_14"  
 /tissue\_type="renal cell adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: kidney; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 222 a 236 c 297 g 180 t  
 ORIGIN

Query Match 86.1%; Score 19.8; DB 152; Length 935;  
 Best Local Similarity 91.3%; Pred. No. 2.8e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccgagccgaagcctggga 23  
 |||||  
 Db 910 CTGCCAGCGGAGGCGCTGGCA 888

## RESULT 2

BF119646/c  
 LOCUS BF119646 959 bp mRNA EST 24-OCT-2000  
 DEFINITION 601757514F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3986834 5', mRNA sequence.

ACCESSION BF119646  
 VERSION BF119646.1 GI:10958790  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 959)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM9192 row: l column: 03  
 High quality sequence stop: 612.

## FEATURES

source  
 Location/Qualifiers  
 1..959  
 /organism="Mus musculus"  
 /strain="C57/B6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3986834"  
 /clone\_lib="NCI\_CGAP\_Mam5"

polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-BJOp library is a subtracted library derived from the UI-R-AAL, UI-R-AAL1, UI-R-AAL1, UI-R-AE1, UI-R-AFL1, and UI-R-AG1 libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV

ACCESSION BF403937  
 VERSION BF403937.1 GI:11391912  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 517)  
 REFERENCE Bernaldo, M.F., Lennon, G. and Soares, M.B.  
 AUTHORS Normalization and subtraction: two approaches to facilitate gene  
 TITLE discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized thalamus library cDNA Library Preparation: M.B. Soares  
 Lab Clone Distribution: clones will be available through Research  
 Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=Yes.

## FEATURES

Location/Qualifiers  
 1..517

/organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone\_lib="R-CAI-bid-j-19-0-UI"  
 /clone\_lib="UI-R-CAI"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAI  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
 , midbrain, cerebral cortex, corpus striatum, testis, and  
 hippocampus. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been  
 previously described in (Bernaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996)  
 TAG\_LIB=UI-R-CAI  
 TAG\_TISSUE=thalamus  
 TAG\_SEQ=ATCGC"

BASE COUNT 114 a 134 c 133 g 136 t  
 ORIGIN

Query Match 84.3%; Score 19.4; DB 148; Length 517;  
 Best Local Similarity 95.2%; Pred. No. 3.9e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gccacggccgaaggcctggga 23  
 ||| ||||| ||||| |||||

Db 313 GCCGAGGCCGGAAGGCGCTGGG 293

## RESULT

BB860673 176 bp mRNA EST 29-SEP-2000  
 LOCUS UI-M-AH1-agv-c-04-0-UI.r1 NIH\_BMAP\_MCE.N Mus musculus cDNA clone  
 DEFINITION UI-M-AH1-agv-c-04-0-UI 5', mRNA sequence.  
 ACCESSION BE860673  
 VERSION BE860673.1 GI:10377856  
 KEYWORDS EST.

## SOURCE

house mouse.  
 Mus musculus

## REFERENCE

1 (bases 1 to 176)  
 AUTHORS Bernaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery

## JOURNAL

Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477

## COMMENT

Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890

Email: mEST@mail.nih.gov

CNA Library Preparation: M.B. Soares Lab Clone Distribution:  
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It  
 should be noted that Bento Soares is generating a small number of  
 additional specialized non-redundant arrays of BMAP cDNAs whose  
 availability will be considered under appropriate and limited  
 collaborative arrangements

Seq primer: M13 Reverse.

## FEATURES

Location/Qualifiers  
 1..176

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="UI-M-AH1-agv-c-04-0-UI"  
 /clone\_lib="NIH\_BMAP\_MCE.N"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_MCE library is a normalized library constructed  
 from mouse cerebellum. The tag is a string of 5  
 nucleotides present between the Not I site and the  
 oligo-dT track. The library was constructed as described  
 by Bernaldo, Lennon and Soares, Genome Research 6: 791-806  
 , 1996. Tissue provided by Ms. Annie Novakovich,  
 Zivic-Miller Laboratories."

BASE COUNT 29 a 58 c 40 g 49 t  
 ORIGIN

Query Match 81.7%; Score 18.8; DB 141; Length 176;  
 Best Local Similarity 90.9%; Pred. No. 6.7e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccacggccgaaggcctggg 22  
 ||| ||||| ||||| |||||

Db 71 CTGCTCAGCCGCGAAGCGCTGGG 92

## RESULT

BB147701 238 bp mRNA EST 28-JUN-2000  
 LOCUS BB147701 RIKEN full-length enriched, adult female vagina Mus  
 DEFINITION musculus cDNA clone 9930101I23 3', mRNA sequence.  
 ACCESSION BB147701  
 VERSION BB147701.1 GI:8802638  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 238)  
 REFERENCE Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
 , P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.







```

plate: L0228 row: C column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 455
POLYA=Yes.
Location/Qualifiers
1. .455
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L0228C11"
/clone_lib="Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/note="vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an Oligo(dT) primer [NotI
primer-adaptor from Gibco/BRL]
[5',_PGAGTAGTTCAGATCGAGCGCCGCTTTT-3'] from
2.5ug of total RNA. The double-stranded cDNAs were
treated with 14 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker L1-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's size fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao."

```

|            |       |       |       |      |
|------------|-------|-------|-------|------|
| BASE COUNT | 109 a | 152 c | 158 g | 36 t |
| ORIGIN     |       |       |       |      |

```
Query Match      80.0%; Score 18.4; DB 117; Length 455;
Best Local Similarity 95.0%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 3 gccacggcgaagcctggg 22  
||||||| |||||||||  
Db 265 GCCACGGCTGAAGGCCTGGG 284

|            |  |
|------------|--|
| RESULT     | 13   |
| BB573895   |  |
| LOCUS      |  |
| DEFINITION |  |
| BB573895   | 286 bp mRNA EST  |
| BB573895   | Riken full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4033400M04 5' mRNA sequence. |
| BB573895   | 30-NOV-2000  |

|           |  |             |
|-----------|--|-------------|
| ACCESSION | BB573895.1   | GI:11470439 |
| VERSION   | BB573895.1   |             |
| KEYWORDS  | EST/73895.1  |             |
| SOURCE    | house mouse  |             |
| ORGANISM  | Mus musculus   |             |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 286) |             |

**AUTHORS**  
Aizawa, K., Kahhira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, T., Hanaoka, T., Hayatsu, N., Hiroaka, T., Hirozane, T., Hodojima, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno, H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takanashi, F., Tanaka, T., Toya, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Aizawa, K. et al. 2000)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki

**TITLE**  
JOURNAL  
COMMENT

ORGANISM      Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 438)  
Hillier, L., Allen, M., Bowles, L., Dubucq, T., Geisel, G., Jost, S.,  
Krisman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin  
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST project  
unpublished (1997)  
Contact: Willson, RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 2561      Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 392.

```

FEATURES
source
Location/Qualifiers
1. .438
/organism="Homo sapiens"
/db_xref="GDB:4622623"
/db_xref="taxon:9606"
/clone="IMAGE:591237"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"
98 a 150 c 115 a 67 t 8 others

```

|             |      |       |       |             |                   |
|-------------|------|-------|-------|-------------|-------------------|
| BASE COUNT  | 98 a | 150 c | 115 g | 67 t        | 8 others          |
| ORIGIN      |      |       |       |             |                   |
| Query Match |      |       | 80.0% | Score 18.4; | DB 3; Length 438; |

|    |     |                        |     |  |  |  |
|----|-----|------------------------|-----|--|--|--|
| Qy | 1   | ctgccccaggccgaaggcctgg | 21  |  |  |  |
|    |     |                        |     |  |  |  |
| Db | 263 | CNCCCCAGGCGGAAGACCTGG  | 243 |  |  |  |

Best Local Similarity 90.5%; Pred.No. le+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

| RESULT     | 12 | EST    | 31-AUG-2000             |
|------------|----|--------|-------------------------|
| AW553523   |    | 455 bp | mus musculus cDNA clone |
| LOCUS      |    |        |                         |
| DEFINITION |    |        |                         |
| ACCESSION  |    |        |                         |
| VERSION    |    |        |                         |
| KEYWORDS   |    |        |                         |
| SOURCE     |    |        |                         |

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 455)

Tanaka, T. S., Jaradat, S. A., Lim, M. K., Kargul, G. J., Wang, X., Grahovac,  
M. J., Pantano, S., Sano, Y., Piao, Y., Nagara, R., Doi, H., Wood, W. H.  
III, Becker, K. G. and Ko, M. S. H.

REFERENCE

AUTHORS

|         |   |
|---------|---|
| TITLE   | Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray  |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)   |
| MEDLINE | 20381348  |
| COMMENT | Contact: George J. Kargul<br>Laboratory of Genetics<br>National Institute on Aging/National Institutes of Health<br>333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA<br>Email: cdna@igsun.qrc.nia.nih.gov |

Genome Exploration Research Group, Life Science Tsukuba Center,  
Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic  
Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rtc.riken.go.jp,  
URL: http://genome.rtc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source  
Location/Qualifiers  
1..286  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="493340M04"  
/clone\_lib="RIKEN full-length enriched, adult male testis  
(DH10B)"  
/sex="male"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI: cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCCC 3']. cDNA  
was cloned into the XhoI and BamHI sites. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."  
BASE COUNT 62 a 74 c 94 g 56 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 161; Length 286;  
Best Local Similarity 87.0%; Pred. No. 1.2e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ctgcccagccgaagcctggga 23  
||||| ||| |||| |||||  
Db 157 CTGCCCGGGCGGAGAGCCTGGGA 179

RESULT 14  
LOCUS BE757761/c 317 bp mRNA EST 15-SEP-2000  
DEFINITION 212291 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE757761  
VERSION BE757761.1 GI:10171753  
KEYWORDS EST.  
SOURCE csw.

## ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

## REFERENCE

## AUTHORS

1 (bases 1 to 317)  
Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,  
Bennett, G.A., Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Laegreid,  
W.W. and Keele, J.W.

## TITLE

Design and use of four pooled tissue normalized cDNA libraries for  
EST discovery in cattle

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390

Email: [smith@email.marc.usda.gov](mailto:smith@email.marc.usda.gov)

Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.

## PCR Primers

FORWARD: AGGAACACAGCTATGACCAT

BACKWARD: GTTTTCCAGTCACGAGC

Plate: 65 row: N column: 2

Seq primer: ATTAGTGCACATATAG.

## FEATURES

## source

1..317  
Location/Qualifiers

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 2BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;  
Library made from pooled tissue from testis, thymus,  
semitendinosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."

## BASE COUNT

64 a 60 c 77 g 116 t  
ORIGIN

## Query Match

79.1%; Score 18.2; DB 139; Length 317;

Best Local Similarity 87.0%; Pred. No. 1.2e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

## QY

1 ctgcccagccgaagcctggga 23

## Db

87 CTGCTCAGGCTCAGGCTGGGA 65

## RESULT

15

## LOCUS

AZ287293

## DEFINITION

RPCI-23-144118.TJ RPCI-23 Mus musculus genomic clone RPCI-23-144118

## ACCESSION

AZ287293

## VERSION

AZ287293.1 GI:9529079

## KEYWORDS

GSS.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 318)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akintet,  
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other\_GSSs: RPCI-23-144118.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Thu Oct 25 13:09:06 2001

us-09-462-955-6.rst

Pax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 144 row: 1 column: 18

Seq primer: SP6  
 Class: BAC ends.  
 Location/Qualifiers  
 1. .318  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-144118"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 89 a 56 c 99 g 74 t

BASE COUNT  
 ORIGIN

Query Match 79.1%; Score 18.2; DB 241; Length 318;  
 Best Local Similarity 87.0%; Pred. No. 1.2e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 ctgcccaggccgagcctggaga 23  
 |||||  
 Db 92 CTGCCAGGCGAGAGACCTTGAGA 114

Search completed: October 24, 2001, 13:14:03  
 Job time: 11828 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:58 ; Search time 5701.85 Seconds  
(without alignments)  
62.393 Million cell updates/sec

Title: US-09-462-955-6  
Perfect score: 23

Sequence: 1 ctgccaggccgaaggctggga 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_bal.\*  
2: gb\_bal.\*  
3: gb\_bal.\*  
4: gb\_bal.\*  
5: gb\_bal.\*  
6: gb\_bal.\*  
7: gb\_bal.\*  
8: gb\_bal.\*  
9: gb\_bal.\*  
10: gb\_bal.\*  
11: gb\_bal.\*  
12: gb\_bal.\*  
13: gb\_bal.\*  
14: gb\_bal.\*  
15: gb\_bal.\*  
16: gb\_bal.\*  
17: gb\_bal.\*  
18: gb\_bal.\*  
19: gb\_bal.\*  
20: gb\_bal.\*  
21: gb\_bal.\*  
22: gb\_bal.\*  
23: gb\_bal.\*  
24: gb\_bal.\*  
25: gb\_bal.\*  
26: gb\_bal.\*  
27: gb\_bal.\*  
28: gb\_bal.\*  
29: gb\_bal.\*  
30: gb\_bal.\*  
31: gb\_bal.\*  
32: gb\_bal.\*  
33: gb\_bal.\*  
34: gb\_bal.\*  
35: gb\_bal.\*  
36: gb\_bal.\*  
37: gb\_bal.\*  
38: gb\_bal.\*  
39: gb\_bal.\*  
40: gb\_bal.\*  
41: gb\_bal.\*  
42: gb\_bal.\*  
43: gb\_bal.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vl.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vl1.\*  
59: gb\_vl2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rod.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_bal.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 23    | 100.0       | 1291   | 58    | CFDCG              |
| c          | 19.8  | 86.1        | 25010  | 90    | M29963 Coconut fol |
| 3          | 19.8  | 86.1        | 185002 | 79    | AL353092 Human DNA |
| c          | 19.8  | 86.1        | 193168 | 62    | AL161784 Homo sapi |
| 5          | 19.8  | 86.1        | 204881 | 81    | AC011969 Homo sapi |
| c          | 19.8  | 86.1        | 207392 | 60    | AL513122 Homo sapi |
| 7          | 19    | 82.6        | 15982  | 1     | AC006558 Homo sapi |
| 8          | 19    | 82.6        | 169194 | 74    | AE004779 Pseudomon |
|            |       |             |        |       | AC073464 Homo sapi |

TRPGVKRPLAQRFAEEDDELRLDPGGYRRCVVGASVETRWAAENPFFPHNQ  
LEVSATCEPADDRITILWICRGDGGKSVFAYKILGLKPDWFTYTCGTRKDLVYQLE  
DKPKNLILDPVRCNLEYLNVALLECKVNRFASSDKYEPILYGLGFDHVLVFNVLDP  
YLKTSRDRILKLNNT  
314. .775  
/note="ORF 2"  
/codon\_start=1  
/protein\_id="AAA42895.1"  
/db\_xref="GI:323308"  
/translation="MTGFTWSRPVVPTRNRIETTVRRNGCFSSSTESRLVLESKGDWPN  
DLRLNMSAWKTOADTEDALYTELWNGDGPGLKIRSHFHTIGSLKCLCLSESORT  
IAQSSGVADETEETGSPCLNISPSTGTHVVEPERTYCTSTSRQNEI"  
complement(422. .568)  
/note="ORF 6"  
/codon\_start=1  
/protein\_id="AAA42896.1"  
/db\_xref="GI:323309"  
/translation="MEMGTDFQRILSTIPPKLVRQIFGRLGLPGGVHQPQIVGP  
IVAF"  
639. .797  
/note="ORF 3"  
/codon\_start=1  
/protein\_id="AAA42897.1"  
/db\_xref="GI:323310"  
/translation="MRTRRRREVVCQISRTQARLVLLHMMWQKGRIVPVHGRPKTK  
FNPRCTQV"  
complement(823. .987)  
/note="ORF 5"  
/codon\_start=1  
/protein\_id="AAA42898.1"  
/db\_xref="GI:323311"  
/translation="WHTLNLIPQFYSVPADFOIIRQDICKYEMHVEPKITKGFVVF  
RTECPVLNTF"  
1098. .1286  
/note="ORF 4"  
/codon\_start=1  
/protein\_id="AAA42899.1"  
/db\_xref="GI:323312"  
/translation="MNRVMGPTIKDSIWRINLLCLCQTOPLSTSIQVSSLLLEKK  
AASLYLPSICFCAIGRLS"  
BASE COUNT 336 a 323 c 332 g 300 t  
ORIGIN

Query Match 100.0%; Score 23; DB 58; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ctgccagccgaagcctggga 23  
Db 1045 CTGCCAGGCCGAGGCGCTGGGA 1067  
RESULT 2  
AL353092/c DNA PRI 11-FEB-2001  
LOCUS  
DEFINITION Human DNA sequence from clone RP1-180113 on chromosome 20 Contains  
5' end of the HCK gene for hemopoietic cell kinase (protein  
tyrosine kinase), contains ESTs, STSS, GSSs and a CpG island,  
complete sequence.  
ACCESSION AL353092  
VERSION AL353092.6 GI:9650539  
KEYWORDS HTG; CpG island; HCK; tyrosine kinase.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 25010)  
Almeida, J.  
Direct Submission  
Submitted (08-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

AL137070 Human DNA  
AF151965 Pseudomon  
AF153207 Pseudomon  
AL117322 Streptomy  
AP001078 Homo sapi  
AP001793 Homo sapi  
AC015676 Homo sapi  
AP000899 Homo sapi  
AF323675 Bradyrhiz  
M13191 Rat tissue-  
U13668 Human G pro  
L32831 Homo sapien  
X83956 H sapiens A  
AK026727 Homo sapi  
M23697 Rat tissue-  
A19618 Mammalian c  
Y11525 H sapiens m  
U18550 Human GPR3  
AJ010482 Homo sapi  
U67167 Homo sapien  
U68061 Human MUC2  
AF016420 Caenorhab  
AL133343 Human DNA  
AL096774 Human DNA  
AL359374 Homo sapi  
AC011651 Homo sapi  
AL078603 Human DNA  
AC021287 Homo sapi  
AL359742 Homo sapi  
AC041043 Homo sapi  
AC024535 Homo sapi  
AL358236 Homo sapi  
AC011935 Homo sapi  
AL590022 Homo sapi  
AC016992 Homo sapi  
AC023331 Homo sapi  
AC016725 Homo sapi

ALIGNMENTS  
CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993  
LOCUS Coconut foliar decay virus, complete genome.  
DEFINITION M29963  
ACCESSION M29963.1 GI:323306  
VERSION circular; complete genome.  
KEYWORDS Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.  
SOURCE Coconut foliar decay virus  
ORGANISM Viruses; ssDNA viruses; Nanovirus.  
REFERENCE 1 (bases 1 to 1291)  
Rohde, W., Randles, J.W., Langridge, P. and Hanold, D.  
Nucleotide sequence of a circular single-stranded DNA associated  
with coconut foliar decay virus  
JOURNAL Virology 176, 648-651 (1990)  
MEDLINE 90266484  
COMMENT Draft entry and printed sequence for [1] kindly submitted by  
W.Rhode, 15-MAR-1989, for release after publication.  
FEATURES  
source  
1. .1291  
/organism="Coconut foliar decay virus"  
/db\_xref="taxon:12474"  
40..70  
/note="stem-loop structure"  
103..975  
/note="ORF 1"  
/codon\_start=1  
/protein\_id="AAA42894.1"  
/db\_xref="GI:323307"  
/translation="MGSSTRMWCFTLVNTEFEAAVVRRIEELNLYVAIVGDEVAPS  
TGQRHLQGFILHLKTRRUGLKVGLNDRIHLEPTRGSDQNRDVCSEKRVLLHGVF

## COMMENT

On Aug 1, 2000 this sequence version replaced gi:9187765. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlap in the sequence. Only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep/](http://www.sanger.ac.uk/projects/C_elegans/wormpep/) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP1-180113. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-836N17 is at 24911 in this sequence. The true right end of clone RP1-310013 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP1-180113 is from the library RPC1-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

## FEATURES

|               |  |                     |  |
|---------------|--|---------------------|--|
| repeat_region | 1. .25010  | Location/Qualifiers |  |
|               | /organism="Homo sapiens"   |                     |  |
|               | /db_xref="taxon:9606"  |                     |  |
|               | /chromosome="20"   |                     |  |
|               | /clone="RP1-180113"  |                     |  |
|               | /clone_lib="RPC1-1"  |                     |  |
|               | 1. .43   |                     |  |
| repeat_region | /note="AluJo/FLAM repeat: matches 76. .118 of consensus"   |                     |  |
| misc_feature  | complement(39. .567)   |                     |  |
| repeat_region | /note="match: GSS: Em:AQ386884"  |                     |  |
| repeat_region | 58. .175   |                     |  |
| repeat_region | /note="5S repeat: matches 1. .119 of consensus"  |                     |  |
| repeat_region | 265. .455  |                     |  |
| repeat_region | /note="MER20 repeat: matches 26. .218 of consensus"  |                     |  |
| repeat_region | 510. .586  |                     |  |
| misc_feature  | /note="L2 repeat: matches 1966. .2045 of consensus"  |                     |  |
| repeat_region | 692. .906  |                     |  |
| repeat_region | /note="match: GSS: Em:AQ486790"  |                     |  |
| repeat_region | 772. .843  |                     |  |
| misc_feature  | /note="2 copies 36 mer 93% conserved"  |                     |  |
| misc_feature  | 1576. .2453  |                     |  |
|               | /note="CpG island"   |                     |  |
|               | /evidence=not_experimental   |                     |  |
| mRNA          | join(1846. .1956,21132. .21252,22789. .22831,23186. .23288,24093. .>24191)   |                     |  |
|               | /gene="HCK"  |                     |  |
|               | /note="match: cDNAs: Em:M16591 Em:X62345 Em:S74141 Em:J03023 Em:M16592 Em:M83666 Em:Y00487 Em:X60380 Em:J03579 Em:X67786 Em:AF000300 Em:AF000301 Em:AF000302 Em:M17031 Em:X52822 Em:X57018 Em:X15345 Em:M27454 Em:M57290 Em:X57684 Em:M85043 Em:M19722 match: ESTs: Em:AW307786" |                     |  |
|               | /evidence=not_experimental   |                     |  |
| gene          | /product="dJ180113.1 (hemopoietic cell kinase)"  |                     |  |
|               | 1846. .24191   |                     |  |
|               | /gene="HCK"  |                     |  |
| misc_feature  | complement(2102. .2519)  |                     |  |
| misc_feature  | /note="match: GSS: Em:AQ382576"  |                     |  |
|               | complement(2108. .2557)  |                     |  |
|               | /note="match: GSS: Em:AQ461389"  |                     |  |
|               | complement(2196. .2570)  |                     |  |
|               | /note="match: GSS: Em:AQ240986"  |                     |  |
|               | 2546. .2600  |                     |  |
|               | /note="MIR repeat: matches 89. .148 of consensus"  |                     |  |
| misc_feature  | 2575. .3113  |                     |  |
|               | /gene="HCK"  |                     |  |
|               | /note="match: GSS: Em:AQ201482"  |                     |  |
|               | 2576. .3351  |                     |  |
|               | /gene="HCK"  |                     |  |
|               | /note="match: GSS: Em:AQ748323"  |                     |  |
| misc_feature  | 2594. .2974  |                     |  |
|               | /gene="HCK"  |                     |  |
|               | /note="match: GSS: Em:AQ821722"  |                     |  |
| repeat_region | 2700. .2825  |                     |  |
| repeat_region | /note="MIR repeat: matches 76. .193 of consensus"  |                     |  |
| repeat_region | 2826. .3131  |                     |  |
| repeat_region | /note="AluSx repeat: matches 1. .306 of consensus"   |                     |  |
| repeat_region | 3132. .3189  |                     |  |
| repeat_region | /note="MIR repeat: matches 193. .252 of consensus"   |                     |  |
| repeat_region | 3221. .3389  |                     |  |
| repeat_region | /note="MERSA repeat: matches 1. .184 of consensus"   |                     |  |
| repeat_region | 3809. .3976  |                     |  |
| repeat_region | /note="AluSp repeat: matches 147. .313 of consensus"   |                     |  |
| repeat_region | 3997. .4296  |                     |  |
| repeat_region | /note="AluY repeat: matches 3. .302 of consensus"  |                     |  |
| repeat_region | 4866. .5115  |                     |  |
| repeat_region | /note="L1MC1 repeat: matches 5725. .5980 of consensus"   |                     |  |
| repeat_region | 5124. .5701  |                     |  |
| repeat_region | /note="L1MB5 repeat: matches 5615. .6171 of consensus"   |                     |  |
| repeat_region | 5702. .5999  |                     |  |
| repeat_region | /note="AluY repeat: matches 1. .295 of consensus"  |                     |  |
| repeat_region | 6000. .6157  |                     |  |
| repeat_region | /note="L1MB5 repeat: matches 5454. .5615 of consensus"   |                     |  |
| repeat_region | 6276. .6378  |                     |  |
| repeat_region | /note="L2 repeat: matches 2623. .2730 of consensus"  |                     |  |
| repeat_region | 6789. .7086  |                     |  |
| repeat_region | /note="AluJb repeat: matches 1. .299 of consensus"   |                     |  |
| repeat_region | 7087. .7147  |                     |  |
| repeat_region | /note="MIR repeat: matches 68. .133 of consensus"  |                     |  |
| repeat_region | 7614. .7706  |                     |  |
| repeat_region | /note="MIR repeat: matches 47. .140 of consensus"  |                     |  |
| repeat_region | 7720. .8014  |                     |  |
| repeat_region | /note="AluSq repeat: matches 1. .297 of consensus"   |                     |  |
| misc_feature  | 7895. .8073  |                     |  |
|               | /gene="HCK"  |                     |  |
| misc_feature  | /note="match: STS: Em:HS336XG5"  |                     |  |
|               | 8076. .8206  |                     |  |
|               | /gene="HCK"  |                     |  |
|               | /note="match: STS: Em:HS336XG5"  |                     |  |
| repeat_region | 8076. .8119  |                     |  |
| repeat_region | /note="22 copies 2 mer ac 97% conserved"   |                     |  |
| repeat_region | 8139. .8510  |                     |  |
| repeat_region | /note="LTR16C repeat: matches 1. .387 of consensus"  |                     |  |
| repeat_region | 8750. .8941  |                     |  |
| repeat_region | /note="48 copies 4 mer cttc 62% conserved"   |                     |  |
| repeat_region | 8779. .8928  |                     |  |
| repeat_region | /note="75 copies 2 mer tt 60% conserved"   |                     |  |
| repeat_region | 8795. .8974  |                     |  |
| repeat_region | /note="5 copies 36 mer 68% conserved"  |                     |  |
| repeat_region | 8820. .8957  |                     |  |
| repeat_region | /note="46 copies 3 mer cct 66% conserved"  |                     |  |
| repeat_region | 8982. .9285  |                     |  |
| repeat_region | /note="AluJo repeat: matches 1. .304 of consensus"   |                     |  |
| repeat_region | 9680. .10012   |                     |  |
| repeat_region | /note="AluSx repeat: matches 1. .311 of consensus"   |                     |  |
| repeat_region | 10208. .10442  |                     |  |
| repeat_region | /note="LTR16B repeat: matches 250. .459 of consensus"  |                     |  |
| repeat_region | 10443. .10751  |                     |  |
| repeat_region | /note="AluSg repeat: matches 1. .309 of consensus"   |                     |  |
| repeat_region | 10752. .10991  |                     |  |
| repeat_region | /note="LTR16B repeat: matches 2. .250 of consensus"  |                     |  |
| repeat_region | 11000. .11406  |                     |  |
| repeat_region | /note="L2 repeat: matches 2157. .2577 of consensus"  |                     |  |





```

/note="assembly_fragment:01103
fragment_chain:3"
102311..175449
/note="assembly_fragment:00558
fragment_chain:3"
175550..177644
/note="assembly_fragment:00056"
177745..182725
/note="assembly_fragment:00871"
182826..185002
/note="assembly_fragment:02179"
53278 a 41233 c 39095 g 50291 t 1105 others
ORIGIN

```

```

Query Match      86.1%; Score 19.8; DB 79; Length 185002;
Best Local Similarity .91.3%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagccgagcctggga 23
Db 136974 CTGCCAGCCAAAGGCTGCAA 136996

```

```

RESULT 4
AC011969/c
LOCUS      AC011969      193168 bp      DNA      HTG      21-APR-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-520J8 map 4, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC011969
VERSION    AC011969.3 GI:7630868
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckhagter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Deaerellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGuck,K., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
            Direct Submission

```

```

TITLE      Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
JOURNAL    Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Apr 21, 2000 this sequence version replaced gi:7008795.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L2649
            Center clone name: 520_J8
            ----- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Assembly program: Phrap; version 0.960731
Consensus quality: 126832 bases at least Q40
Consensus quality: 155745 bases at least Q30
Consensus quality: 175228 bases at least Q20
Insert size: 186000; agarose-fp
Insert size: 191968; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; agarose-fp
Quality coverage: 3.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as-soon as it is available and the accession number will
* be preserved.

```

```

* 1 1093: contig of 1093 bp in length
* 1094 1193: gap of 100 bp
* 1194 2376: contig of 1183 bp in length
* 2377 2476: gap of 100 bp
* 2477 5933: contig of 3457 bp in length
* 5934 6033: gap of 100 bp
* 6034 9960: contig of 3927 bp in length
* 9961 10060: gap of 100 bp
* 10061 16555: contig of 6495 bp in length
* 16556 16655: gap of 100 bp
* 16656 22671: contig of 6016 bp in length
* 22672 22771: gap of 100 bp
* 22772 31113: contig of 8342 bp in length
* 31114 31213: gap of 100 bp
* 31214 41484: contig of 10271 bp in length
* 41485 50824: contig of 9240 bp in length
* 50825 50924: gap of 100 bp
* 50925 69485: contig of 18561 bp in length
* 69486 69585: gap of 100 bp
* 69586 99210: contig of 29625 bp in length
* 99211 99310: gap of 100 bp
* 99311 130378: contig of 31068 bp in length
* 130379 130478: gap of 100 bp
* 130479 193168: contig of 62690 bp in length.

```

## FEATURES

## Source

```

1. 193168
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-520J8"
/clone_lib="RPC1-11 Human Male BAC"
1. 1093
/note="assembly_fragment"
1194. 2376
/note="assembly_fragment"
2477. 5933
/note="assembly_fragment"
6034. 9960
/note="assembly_fragment"
10061. 16555
/note="assembly_fragment"
16656. 22671
/note="assembly_fragment"
22772. 31113
/note="assembly_fragment"
31214. 41484
/note="assembly_fragment"
41585. 50824
/note="assembly_fragment"
50925. 69485
/note="assembly_fragment"
69586. 99210
/note="assembly_fragment"
99311. 130378

```

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

```

/notes="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature
130479..193168
/notes="assembly_fragment"
BASE COUNT 55031 a 40516 c 42876 g 53468 t 1277 others
ORIGIN

Query Match 86.1%; Score 19.8; DB 62; Length 193168;
Best Local Similarity 91.3%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccagccgcaaggcctggga 23
|||||
Db 115095 CTGCCAGCCCAAGGCTGGAA 115073

RESULT 5
AL513122 204881 bp DNA HTG 25-FEB-2001
LOCUS Homo sapiens chromosome 9 clone RP11-477J21, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.
ACCESSION AL513122
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Pavitt,R.
Direct Submission
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 28, 2001 this sequence version replaced gi:12578283.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA477J21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 201478 bases at least Q40
Consensus quality: 202815 bases at least Q30
Consensus quality: 203621 bases at least Q20
Insert size: 204181; sum-of-contigs
Insert size: 206137; 6.5% error; agarose-fp
Quality coverage: 6.47x in Q20 bases; sum-of-contigs Quality
coverage: 6.41x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 15843 15843: contig of 15843 bp in length
* 15944 15944: gap of 100 bp
* 36523 36523: contig of 20580 bp in length
* 36624 36624: gap of 100 bp
* 67681 67681: contig of 31058 bp in length
* 67782 67782: gap of 100 bp
* 119925 119925: contig of 52144 bp in length
* 119926 120025: gap of 100 bp
* 120026 120026: contig of 42664 bp in length

```

```

* 162690 162789: gap of 100 bp
* 182790 174915: contig of 12126 bp in length
* 174916 175015: gap of 100 bp
* 175016 193301: contig of 18286 bp in length
* 193302 193401: gap of 100 bp
* 193402 204881: contig of 11480 bp in length.
FEATURES
Location/Qualifiers
source
1..204881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-477J21"
/clone_lib="RPC1-11.2"
1..15843
/notes="assembly_fragment:00222"
clone_end:17
vector_side:left
15944..36523
/notes="assembly_fragment:02927"
36624..67681
/notes="assembly_fragment:00442"
fragment_chain:1
67782..119925
/notes="assembly_fragment:02072"
fragment_chain:1
120026..162689
/notes="assembly_fragment:03061"
fragment_chain:1
162790..174915
/notes="assembly_fragment:00959"
fragment_chain:1
175016..193301
/notes="assembly_fragment:01787"
fragment_chain:1
193402..204881
/notes="assembly_fragment:01691"
fragment_chain:1
clone_end:SP6
vector_side:right
57526 a 45247 c 44379 g 57028 t 701 others
ORIGIN

Query Match 86.1%; Score 19.8; DB 81; Length 204881;
Best Local Similarity 91.3%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccagccgcaaggcctggga 23
|||||
Db 39729 CTGCCAGCCCAAGGCTGGAA 39751

RESULT 6
AC006558/c 207392 bp DNA HTG 10-FEB-1999
LOCUS Homo sapiens chromosome 9 clone hRPK.477_J_21 map 9, *** SEQUENCING
DEFINITION IN PROGRESS ***, 7 unordered pieces.
ACCESSION AC006558
VERSION AC006558.1 GI:4262259
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 207392)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 9, clone hRPK.477_J_21
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207392)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,

```

TITLE  
JOURNAL  
COMMENT

Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McCurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Nilloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission  
Submitted (10-FEB-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3313: contig of 3313 bp in length  
\* gap of unknown length  
\* 3314 9632: contig of 6319 bp in length  
\* gap of unknown length  
\* 9633 24168: contig of 14536 bp in length  
\* gap of unknown length  
\* 24169 38366: contig of 14198 bp in length  
\* gap of unknown length  
\* 38367 56172: contig of 17806 bp in length  
\* gap of unknown length  
\* 56173 87103: contig of 30931 bp in length  
\* gap of unknown length  
\* 87104 207392: contig of 120289 bp in length.

## FEATURES

Location/Qualifiers  
1..207392  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hrpk-477-J-21"  
/clone.lib="RPC1-II human BAC library"  
/map="g"  
/chromosome="9"  
BASE COUNT 57991 a 44607 c 45345 g 58166 t 1283 others  
ORIGIN

Query Match 86.1%; Score 19.8; DB 60; Length 207392;  
Best Local Similarity 91.3%; Pred. No. 68;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 ctgccagcggcgaagcctggga 23  
||||||| |||||  
Db 70048 CTGCCAGGCCAAGCCTGGAA 70026

RESULT 7  
AE004779  
LOCUS  
DEFINITION Pseudomonas aeruginosa PA01, section 340 of the complete genome, BCT 30-AUG-2000  
ACCESSION AE004779 AE004091  
VERSION AE004779.1 GI:9949735  
KEYWORDS  
SOURCE Pseudomonas aeruginosa.  
ORGANISM Pseudomonas aeruginosa.  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.  
REFERENCE  
1 (bases 1 to 15982)  
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,

Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.  
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
Nature 406 (6799), 959-964 (2000)  
20437337  
2 (bases 1 to 15982)  
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

## FEATURES

source

1..15982

/organism="Pseudomonas aeruginosa"

/strain="PA01"

/db\_xref="taxon:287"

163..1002

/gene="glpP"

/note="PA3581"

163..1002

/gene="glpP"

/codon\_start=1

/transl\_table=11

/product="glycerol uptake facilitator protein"

/protein\_id="AAG06969.1"

/db\_xref="GI:9949736"

/translation="MTTAAPTPSLFGCLAEFLGTALLIFFGTCVAALKVAGASPLG

WEISLIMGVGVSMALYLSAGVSGAHLNPAVSLALWLFAGPEGKLPFYITAAQVAGFC

AAALVYLYSLFIEFEQANIVRGSQDSLALASVFSTYPHPALSYQAFLEPVVITTA

ILMAVIMALTDDGNGDLGRPLAPILLIGLIAVIGSAMGLTGFAMPNAPDFGPKLMTY

LAGWGTAFTGGREIPYFLVPIFAPILGACLGAGGYRLIARILPASAAPAEPEKV

RAS"

1042..2559

/gene="glpK"

/note="PA3582"

1042..2559

/gene="glpK"

/codon\_start=1

/transl\_table=11

/product="glycerol kinase"

/protein\_id="AAG06970.1"

/db\_xref="GI:9949737"

/translation="MTDKHNKYYVALDQGTSSRAIVFDRDANVVQAQREFAQFYP

QAQVHDPMEIWAQSSTLVEALQAASIEHQVAAIGITNORETTVVWDRHSGRPIH

NAIVQCRSAAICAOCLKDGLDYIRETTGLVTDYFSGTKLWILONVEGAREAR

NGDLLFCTIDTWLWIKLITEKVHTDYTNASRTMLFNHRSRDWARMLEVLDPKSMIL

PEVRNSEVYGNARIGVGGGELPIAGIAGDQQAALFGOMCEPGAOKNTYTGCGFL

MHTGDKAVKSTHGLITTIACGPRGEVYALEGAVNGSGTVQMLDELKVINDSFDSE

YFAIKVDSNGVYLPFTGLGAPYWDYARGAVFLTRGVKADHLIRATLESIAVQT

RDVLDKADQAGERLARLVDDGAVANNEFLMQFADILCTVRVPMRETTALGAAYL

AGLACGFWSLSLDELKSKAVIERVFECEDEPREKLYAGWKKAVERTRGDDGEL"

2765..3520

/gene="glpP"

/note="PA3583"

2765..3520

/gene="glpP"

/codon\_start=1

/transl\_table=11

/product="glycerol-3-phosphate regulon repressor"

/protein\_id="AAG06971.1"

/db\_xref="GI:9949738"

/translation="MNLPPROQSLIEEVRERYGLSIEEMAQFAVTPOTIRRDINOLA

EGGLRHYHGGAAVSSDIENTAYTMRADMRDEKORIAEAVASLVPPNASILFINCTT

TEAIFARALLNHRNLKIITNNLHVAATLSAKEDFEVLVAGCTVRSDGGIVGQAVDFIQ

```

QFRVDFALVGIISGIDEDGSLDDFDYOEVRVSOAIIDNARQVFLAADSKFGKGNVRL
GPTSLNVRVISOSPPPAALARLMAQKHVLDIV
3798..5336
/gene="g1pD"
/note="PA3584"
3798..5336
/gene="g1pD"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate dehydrogenase"
/protein_id="AAG06972.1"
/db_xref="GI:9949739"
/transl_table="MSQAHTPSAPLAEVYDVAVVGGINGVGIADAGRGSLVFLCE
QHDLAQHTSASSKLIHGGLRYLEHFRVLRALAEVLLAKAPHIVKPLRVLP
RPHLRPAWMTAGLFLYDHLGKREKLPASRGFLTGSPLKAEIRRGFEYSOCVDDA
RLVVLNAISAREGAHVHTTRCVSARRSKGLWHLHLSRSDGSLYIRARALVNAAGP
WVARFIDDDLKQSPYGRILIOGSHLIVPKLYEGEHAYILONEDRIVEAIPYLDORFT
MIGTDREYDDPAKVAISEETAYILQVYNHFKOOLAAADILHSHFAGVRPLCDDES
DEPSAIDTRDYLTLSSAGNEPPLLSVFGGLTKITRKLASALTQLQFPFANLGPWTA
KAPLPGEQMSQVEALTEOLANRYAMLDELARLWARTYGTVRWRLDGVNGEADLGE
HLGGGLYAREVDYLCKHEWAQDAEDILWRRSKLGFLSPSQQVRLGQYLQSEPHPR
VHAA"
5469..5798
/gene="g1pM"
/note="PA3585"
5469..5798
/gene="g1pM"
/codon_start=1
/transl_table=11
/product="membrane protein G1pM"
/protein_id="AAG06973.1"
/db_xref="GI:9949740"
/transl_table="MFKALIGAAVVWLLAVLSKTRNYYIAGLVPLEPTFALIAHYLVG
KGRSLDDKTTIVFGMWSIIPFYFLAALYLLVERFRLETSALAAWLVAASVLVG
LWVRLHA"
complement(5815..6801)
/gene="PA3586"
complement(5815..6801)
/gene="PA3586"
/codon_start=1
/transl_table=11
/product="probable hydrolase"
/protein_id="AAG06974.1"
/db_xref="GI:9949741"
/transl_table="MRAEFLAALICSLSSFAAACDARVPTETVELGDVRLAYQSIG
RDSFALLMGLGQLHWDPEVVSALCEGQFVIRYDNEVGLSANNVVPSSRLT
YEVVRYLGLPWSADPYTLTDWAGDALHLLDALDIPQAVILGASMGHIAQHIADMAQ
RLSLTLVMTSGSAGLPAPESELRLRLARREASREQAVEQADLLAAGSPVPRDD
ROOLLQARSYDRAFNEGVORQLLAIAPESRVPPLNRLQVPTLVIGHGTAPLLPV
MHGVHVAHHSYRSELKLPGLAHRFQEAKEPFLAAVVPYILKAHQRAQSVVOL"
7100..8020
/gene="metr"
/note="PA3587"
7100..8020
/gene="metr"
/codon_start=1
/transl_table=11
/product="transcriptional regulator Metr"
/protein_id="AAG06975.1"
/db_xref="GI:9949742"
/transl_table="MLELRHLKTLHALREADSLVAAERLHLTQSALSHQPKLEERL
GMOLFVRKTPVRFTSAGLRLQLADSLLPOLRTAERDLARLGGTAGRLHMALECHS
CFOWLMPTIDQFDAMPEVELDLASGFSAPLAPALAGDLVVTADVEVLPFTYVP
LFTYVALLAVANOHLAGRPYVVPEDLERETLITYPVERDRLDVFTFRLDPADVEPAQ
VRIAEITYMMQLVASGRCVCCLPNALHEYSRGVYVITARRLGERGLYCTLFAAIRTD
MEDAPFMRDPLLTAKSDITSPATLEGVSAAG"
complement(8026..9276)
/gene="PA3588"
complement(8026..9276)
/gene="PA3588"
/codon_start=1
/transl_table=11
/product="probable porin"

```

```

/protein_id="AAG06976.1"
/db_xref="GI:9949743"
/transl_table="MNNKCAFEFLPILAGDSVPALAEGLFEDSRASLALRNFYMR
DFRAGRAKSEWAQGFLEFDRSYGTEGLGVGLDLGLKLVGLVRLDSGASGTGLLP
LRDDGSAAGYARLDATLRSSELKLVGLVLPKPLPTIOPNVRGLFPPOVFOGALLTS
GELSCLSLNGLRTEVSORNEAGTSDIALFNRRNRFAQAADRFLDGLADYRIAPDW
TGSYHGELEQVYAHLEGLKIGRIAGDSLESLRLALSRTDGGAGRRGIDNKSFSG
SLTYRLNGOAFGIGYQRMGSDHGFVPLEGTDPLYLVNFQYNDFAEGESSWQIRYDC
DFAPLGVPLSLMTRFESGHCAKPKGDSREWERDSDLRVYVQGGALKGLGIVWRNA
TYRSFASRDIDENRLYLYTELPLF"
complement(9329..10534)
/gene="PA3589"
complement(9329..10534)
/gene="PA3589"
/codon_start=1
/transl_table=11
/product="probable acyl-CoA thiolase"
/protein_id="AAG06977.1"
/db_xref="GI:9949744"
/transl_table="MLDAYIYAGLRTFPRHAGALSTVRPDDLGLALLARLAETSGFA
VDDLEDTLTCNOAGDSRNLARNALLAAGLPAFLPGQTVNRLCASGLSAVIDAARA
ISCGEGRLYAGGAEFSRPAPEVMGKAFSAFRTLEVFDSTIGARFANPRIVRYGND
LIVERDEHPRPQADLAALRLPALFAGGVYFAGNAGSINDGAUVLLGDRAIGREGIR
PLARILASASVGVPEPLMGIGPQQAILRALORAGIDDEVGLIELINEAFAPQVLACLK
LIGLDYEDPRVYNPHGGATLGHPLGASGARILVLTAAARGLQRIERRIYVSLCVGLGQC
VAMVIERCR"
complement(10557..12089)
/gene="PA3590"
complement(10557..12089)
/gene="PA3590"
/codon_start=1
/transl_table=11
/product="probable hydroxyacyl-CoA dehydrogenase"
/protein_id="AAG06978.1"
/db_xref="GI:9949745"
/transl_table="MSEPSIVLIGIVGTGAGGQIAOLAACAGLSVLLYDSRQGVALQ
AREQIATVLAQVERGRLEAEAVRAMGNLYVEDLVLGGCOLVIEAVENLEAKQA
LFRQLEEVVGGDEAILASNTSSVTAISACRDPGRIGLHFFNPVPMRLIVEVTEGL
ATRTGIAERLCALVATFQHVAVRATDPSGFIVNHAGRAGFTGALRILGEGVAPVAID
EVLREGAGFRMGPFELFDVLGDLVSLPMESYIROYEPRYRPHPLRLQMLAARLIG
RKSGQGFYRYDAGQVPAAPVGAALPPVWLGVDDHDEHRAPIMLMLQRLGAEVES

```

```

Query Match      82.6%; Score 19; DB 1; Length 15982;
Best Local Similarity 100.0%; Pred. No. 4,2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gccacggccgaagcctgg 21
    |||||
Db 8404 GCCACGCCGAGGCCCTGG 8422

RESULT 8
AC073464 AC073464 169194 bp DNA HTG 15-JUL-2000
LOCUS Homo sapiens chromosome 2 clone RP11-134N21, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
AC073464 AC073464 2 GI:8954226
VERSION AC073464.2 GI:8954226
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169194)
AUTHORS The sequence of Homo sapiens clone
TITLE Waterston,R.H.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169194)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (18-JUN-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

MO 63108, USA  
On Jul 7, 2000 this sequence version replaced gi:8571786.

COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0134N21  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: plasmid; 0%  
Chemistry: Dye-terminator ET; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 160139 bases at least Q40  
Consensus quality: 162624 bases at least Q30  
Consensus quality: 163974 bases at least Q20  
Insert size: 166000; agarose-fp  
Quality coverage: 4.34 in Q20 bases; agarose-fp  
Quality coverage: 4.36 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 16 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1364: contig of 1364 bp in length  
\* 1365 1464: gap of unknown length  
\* 1465 3740: contig of 2276 bp in length  
\* 3741 3840: gap of unknown length  
\* 3841 7587: contig of 3747 bp in length  
\* 7588 7687: gap of unknown length  
\* 7688 11136: contig of 3449 bp in length  
\* 11137 11236: gap of unknown length  
\* 11237 14308: contig of 3072 bp in length  
\* 14309 14408: gap of unknown length  
\* 14409 21712: contig of 7304 bp in length  
\* 21713 21812: gap of unknown length  
\* 21813 27149: contig of 5337 bp in length  
\* 27150 27249: gap of unknown length  
\* 27250 35937: contig of 8688 bp in length  
\* 35938 36037: gap of unknown length  
\* 36038 47252: contig of 11215 bp in length  
\* 47253 47352: gap of unknown length  
\* 47353 58329: contig of 10977 bp in length  
\* 58330 58429: gap of unknown length  
\* 58430 72793: contig of 14364 bp in length  
\* 72794 72893: gap of unknown length  
\* 72894 85444: contig of 12551 bp in length  
\* 85445 103459: gap of unknown length  
\* 103460 103559: gap of unknown length  
\* 103560 124591: contig of 21032 bp in length  
\* 124592 124691: gap of unknown length  
\* 124692 147042: contig of 22351 bp in length  
\* 147043 147142: gap of unknown length  
\* 147143 169194: contig of 22052 bp in length.  
Location/Qualifiers  
1. .169194  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-134N21"  
1. .1364  
/note="assembly\_name:Contig4"  
1465. .3740  
/note="assembly\_name:Contig5"  
3841. .7587

FEATURES  
source

misc\_feature  
misc\_feature  
misc\_feature

misc\_feature /note="assembly\_name:Contig6"  
7688. .11136  
misc\_feature /note="assembly\_name:Contig7"  
11237. .14308  
misc\_feature /note="assembly\_name:Contig8"  
14409. .21712  
misc\_feature /note="assembly\_name:Contig9"  
21813. .27149  
misc\_feature /note="assembly\_name:Contig10"  
27250. .35937  
misc\_feature /note="assembly\_name:Contig11"  
36038. .47252  
misc\_feature /note="assembly\_name:Contig12"  
47353. .58329  
misc\_feature /note="assembly\_name:Contig13"  
58430. .72793  
misc\_feature /note="assembly\_name:Contig14"  
72894. .85444  
misc\_feature /note="assembly\_name:Contig15"  
clone\_end:17  
vector\_side:right  
85545. .103459  
misc\_feature /note="assembly\_name:Contig16"  
103560. .124591  
misc\_feature /note="assembly\_name:Contig17"  
124692. .147042  
misc\_feature /note="assembly\_name:Contig18"  
147143. .169194  
/note="assembly\_name:Contig19"  
BASE COUNT 51420 a 32271 c 33139 g 50860 t 1504 others  
ORIGIN

Query Match 82.6%; Score 19; DB 74; Length 169194;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gccacggccgaaggcctgg 21  
|||||  
Db 43409 GCCACGGCCGAAGGCGCTGG 43427

RESULT 9  
ALI37070/c  
LOCUS ALI37070 170237 bp DNA PRI 27-JUL-2000  
DEFINITION Human DNA sequence from clone RP11-251017 on chromosome 9, complete  
sequence.  
ACCESSION ALI37070  
VERSION ALI37070.9 GI:9367917  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Skuce,C.  
Direct Submission  
Submitted (27-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Jul 22, 2000 this sequence version replaced gi:8977609.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Embl, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-251017 is from the library RPII-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pBACe3.6 This sequence is the entire insert of clone RP11-251017.

```

FEATURES             source
    1..170237
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="9"
    /clone="RP11-251017"
    /clone_lib="RPII-11.1"
    /note="Random repeat. Single clone region. Assembly confirmed by restriction enzyme digest data."
misc_feature         98259..98301
    /note="Tandem repeat. Short join in tandem repeat confirmed by ECORI and HINDIII digest data."
misc_feature         98302..98403
    /note="Tandem repeat. Single clone region. Assembly confirmed by restriction enzyme digest data."
BASE COUNT          54028 a 32613 c 32057 g 51539 t
ORIGIN

```

```

Query Match          82.6%; Score 19; DB 89; Length 170237;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gccacggccgaaggctgg 21
|||||
DB 26857 GCCACGGCCGAAGGCTGG 26839

```

```

RESULT 10
AF151965 5162 bp DNA BCT 01-JUN-2000
LOCUS Pseudomonas putida strain KT2440 proline dehydrogenase (putA) gene,
DEFINITION complete cds.
ACCESSION AF151965
VERSION AF151965.1 GI:8132051
KEYWORDS Pseudomonas putida.
SOURCE Pseudomonas putida
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 5162)
AUTHORS Vilchez,S., Molina,L., Ramos,C. and Ramos,J.L.
TITLE Proline catabolism by Pseudomonas putida: cloning, characterization and expression of the put genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5162)
AUTHORS Vilchez,S., Molina,L., Ramos,C. and Ramos,J.L.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-1999) Biochemistry, CSIC, Profesor Albareda No.1, Granada E-18008, Spain
FEATURES             Location/Qualifiers
    source            1..5162
                        /organism="Pseudomonas putida"
                        /strain="KT2440"
                        /db_xref="taxon:303"
    gene              359..4306
    CDS               359..4306

```

```

/feature="putA"
/feature="putA; proline oxidase; pyrroline 5'carboxylate
dehydrogenase"
/codon_start=1
/transl_table=11
/product="proline dehydrogenase"
/protein_id="AAF73193.1"
/db_xref="GI:8132052"
/translation="MATTTLVKDDPTREKAAOSIDRPHLLIKQAFVNFLEKL
EGATVTLNGLHASPADDAGEVQADSHQCFLEPAESILPOSVLSAITAAYRRPQ
EYVPMLLQARLAPADATNKLAASTAEKLRNOKSVGGAGIVQGLLOFESSLOER
RGVCLAEALLRIPDKRDALIRKISTGNQPHLGNPSLVFVNAATWGLLTGKLV
THNETGLSSLTIRIKSGEPMIRKVDMAIRMGEOFTVGETTAALANASFEAKG
FRYSYDMLGEALTEHDAQKYLASVEQAIHSIGKASHGRGIEGPGISIKLSALHPRY
SRAQYRMEELYPRLLSILIAKOYDIGINADAEADRLSLDLRLERLCFEFLAG
WNGIGVIAQKRCPYLLNYFEDLAKYPTPHRLIRLLKGYMDESLKRAQVEGLECY
PVYTRKYDVSIVACAKKLLAPYALYQFATHNAHTLSAIYHAGONYPGQYEFQ
CLHGMGEPLYEQVVKIADKLNRCRVAPVGTHTETLLAYLVRRLLGANTSFVNR
IADHSISIQELVADRWPAIAWPKRSGISGLPHRIPRLPOLYGTERRAKLAGIEMANE
HRLGLISCAMVATAHKQWEAPLLACAAASAAPVLPADNRNVVGHQVATVAKFD
NATHCALNPATWQATPPAERAAIILERTADLMEAIHPIMGLLIREAGKTFPNAIEL
REAVDFLYVAQALNDFSNDARPLGPVVCISWNFPLAIFTQVAAALAGNPVLA
KBAQPTLIAQAQVRLLEAGIPGVLQIPLGRGTVCAGLVGDERVGVNFTSTEV
ARLLQNRVAGLDNRGRIPLIAETGGONAMIVOSSALTEQVVIDVSSAFDSAGOR
SALRVLCLEDSDARVIELKMAESRGCDDRLAVDIGIVDAEAKAGIEKHIOGM
REKREYVQVATADAETIKRGTTFVMTLIEDSFDLKRREIFGPVHVYVYRNRNLOQ
LIBQINNSGYGLTGLVHTRIDETIAKVETATPATCRHNRIVGAVGVQPPGGGLSG
TGPKAGGPLYLYRLITRPADAIGRPHQOQDGETPDRTHLHQLVKPLHGLKAWAENN
QADLAALCQFASOSQGIARLLPRELOAKVKLVADNKNKDEFAVDVAVTHHGDSDQLRG
VLAVSSAVVWVDEGPKALRRLPRELOAKVKLVADNKNKDEFAVDVAVTHHGDSDQLRG
VCOQVAKRAGAIVGHVGLSSGDHQTALERLVIERAVSVNTAAAGGNASLMTIG"
BASE COUNT          1051 a 1702 c 1502 g 907 t
ORIGIN

```

```

Query Match          81.7%; Score 18.8; DB 2; Length 5162;
Best Local Similarity 90.9%; Pred. No. 8.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgccacggccgaaggctggga 23
|||||
DB 1681 TGCCACGGCCGAAGGCTGGAA 1702

```

```

RESULT 11
AF153207/c 8660 bp DNA BCT 16-JAN-2000
LOCUS Pseudomonas putida proline dehydrogenase (putA) and proline
DEFINITION permease (putP) genes, complete cds.
ACCESSION AF153207
VERSION AF153207.1 GI:5693828
KEYWORDS Pseudomonas putida.
SOURCE Pseudomonas putida
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 8660)
AUTHORS Vilchez,S., Molina,L., Ramos,C. and Ramos,J.L.
TITLE Proline catabolism by pseudomonas putida: cloning, characterization, and expression of the put genes, in the presence of root exudates
JOURNAL J. Bacteriol. 182 (1), 91-99 (2000)
MEDLINE 20082847
PUBMED 10613867
REFERENCE 2 (bases 1 to 8660)
AUTHORS Vilchez,S., Molina,L., Ramos,C. and Ramos,J.L.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Biochemistry, CSIC, Profesor Albareda No 1, Granada E-18008, Spain
FEATURES             Location/Qualifiers
    source            1..8660
                        /organism="Pseudomonas putida"
                        /strain="KT2440"

```

```

/db_xref="taxon:303"
complement(857..4804)
/gene="putA"
CDS
complement(857..4804)
/gene="putA"
/codon_start=1
/transl_table=11
/product="proline dehydrogenase"
/protein_id="AAF25000.1"
/db_xref="GI:6693829"
/translation="MATTTILGVKLDPTRELRKAAQSIDRTPHWLKQAIIFYLEKLEK
EGGATITELNGHASNADDAEVDHSHQCELEFAESILQSVLRSATIAAYRRPEO
EVPMLLEQARUSPADATNKLAKIAEKNQKIVSGVGRAGIVGLQLEFSLSQOER
RGVCLAEALLRPDKGTRDALIRDKISTGNMOPHLGNSPSLFVNAATWGLLTGKLS
THNETGITSLLYRIIGKSGEPMIRKGVDMMLRMLGEQVETGETAEALANAREAKG
FRYSYDMLEGEALYTHDAOKYLAHYEQAETHSGKASHGSGYEGEGISIKUSALHY
SRAOYERMEELYPRLLSITLAKQYDITGLINDAEADRLESLDLERLEFESLAG
WNGIGVIAQYKRPYLINFFDLAKRTPHRLIIRLLKAYWDSERIKRAQVELEGY
PYTRKVTYDSIVACARKLAVPEALYPOFATHNAHTLSAIYHAGONYPGQYEFQ
CLHMGEPLEYEQVKGADKLNRRCPYAPVGTHTETLLVLRLLENGANTSVNR
IADHSISQELVADRWPAISIAWPKKSGISGLPHRPIPLPRDLYGTERAKLAGIEMNE
HRLGLSCAMVATKOWEAPALLACAARESAAAPVLNPAHRNVVGVHQVQATVAKED
NATHCALNPATWQATPPAERAAIILERTADLMEAEIHPMLGILLIREAKTFPNAIEL
REAVDFLYVAQVALNDSNDHRELGPVVCISPNWFLPFIATFGOVAAALAGNPVLA
KPAEQTPLIAAGARVLLLEAGIEGVQLLIPRGETVAGLVGDERVKGMFTGSTEV
ARLLQNVAGRLDNODGRPTPLIAETGSGONAMIVDSALTQEWIDVYSSAFDSAGORC
SALRLVLCQEDSADRVIEMLGAMAESRLGCPDLAVDIGPVIDAEAKAGIEKHIOGM
REKGRPVQVALADAAEIKRGTFVMTLIELDSDELKREFGVLHVYRNRLDQ
LLEOINNSGYLTLGVHTRIDETAKVVEATPATCRRNRNIVGAVGVQPPGEGELSG
TQPKAGGPLYLVRLSTRPDAIGRHFOQDQEGTDPDRLHEQLKPLHLGKAWAENN
OLADUALCSQFASQSGIARLIPGPTGERSNYTILPREHVLCLADNETDLLAQFAA
VLAVGSSVVDVDEGPKALRPRELOAKYKLVADWNKDEVAFDVFIHGGDSQDLRG
VCOQVAKRAGAIIVGHLSGSDHQIATERLIERASVNTAAAGNLSMTIG"
5160..6639
/gene="putP"
CDS
5160..6638
/gene="putP"
/codon_start=1
/transl_table=11
/product="proline permease"
/protein_id="AAF25001.1"
/db_xref="GI:6693830"
/translation="MGNPLTITFVIYITAMVLTIGFAAYRATNMLSDYILGGRSLGSV
TALSAGSDMSQWLLMLGPGAIYFAGLSEAWIAGTUGVAYLNLVAGRLRVQTEHN
GDALTLPYFSSREDNSGLLRIISAIVILVFETICASGIVAGARLFESTQMSYET
ALWAGAAITAYTGVGLAVSWDTIVQASLMIFALITPVIILSTGGFDQTFAAIE
AVNRNFDLKGATFIGIISLMGWLGFRRPHILARFMAADSVNIAKARRISMTWM
ILCLAGTCAVGCFCSIAVESAHPPELAGPVSFNHKKVFTIELAKILFNPNVAGVLLSALLA
AVMTLSQCLLVCSALTENFYKAFIRKNASQVLYVWGLVLAVALIAIAMAANPE
NRVLGLVAYAWAGFGAAGFPVLLISLVKMGTRNGALAGIIVGVALTIVLWKNFDTLGL
YEIIPGFLFASIAIVLSKLSGSPQTHVKKRFEADAAYHADK"
BASE COUNT 1563 a 2590 c 2748 g 1759 t
ORIGIN

Query Match
Best Local Similarity 81.7%; Score 18.8; DB 2; Length 8660;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcacagccgaaggcctggga 23
||||||| |||||||||
Db 3482 TGCCAGGTCGAAGCGCTGGAA 3461

RESULT 12
LOCUS SCF1 14750 bp DNA BCT 09-SEP-1999
DEFINITION Streptomyces coelicolor cosmid F1.
ACCESSION AL117322
VERSION AL117322.1 GI:5869937
KEYWORDS acetyltransferase; alcohol dehydrogenase; binding-protein-dependent
transport; hydrolase; hydroxylase; LysR-family; MarR family;
oxidoreductase; transcriptional regulator.

```

Streptomyces coelicolor A3(2).  
Streptomyces coelicolor A3(2).  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 14750)  
Redenbach, M., Kieser, H.M., Denapatte, D., Eichner, A., Cullum, J.,  
Kinashi, H. and Hopwood, D.A.  
A set of ordered cosmids and a detailed genetic and physical map  
for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
2 (bases 1 to 14750)  
Seeger, K.J. and Harris, D.  
Unpublished  
3 (bases 1 to 14750)  
Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
Direct Submission  
Submitted (09-SEP-1999) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK  
Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL: http://www.sanger.ac.uk/projects/S.coelicolor/)  
CDS are numbered using the following system eg SC7B7.01c, SC (S.  
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous.  
The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon.  
Gene prediction is based on positional base preference in codons  
using a specially developed Hidden Markov Model (Krogh et al.,  
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot  
program of Bibb et al., Gene 30:157-66(1984) as implemented at  
http://www.nih.gov/jp/  
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or att) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.  
Cosmid F1 Lies between and overlaps with cosmids F20 and F85 on the  
AseI-F genomic restriction fragment.  
Location/Qualifiers  
1..14750  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid F1"  
complement(1..187)  
/note="Nominal overlap with Streptomyces coelicolor cosmid  
F20."  
<1..505  
/gene="SCF1.01"  
/note="SCF1.01, partial CDS, probable alcohol  
dehydrogenase, len: >167aa; similar the C-terminal region  
of many eg. SW:ADH3\_SULSO NAD-dependent alcohol  
dehydrogenase from Sulfolobus solfataricus (347 aa) fasta  
scores; opt: 274, z-score: 304.7, E(): 1.3e-09, (31.8%  
identity in 170 aa overlap). Also similar to C-terminal  
region of TR:BAA82700 (EMBL:AB017438) from Streptomyces

```

coelicolor (346 aa) fasta scores; opt: 372, z-score:
409.3, E(): 2e-15, (40.7% identity in 167 aa overlap).
/codon_start=2
/transl_table=11
/label=SCF1.01
/product="putative alcohol dehydrogenase"
/protein_id="CAB55521.1"
/db_xref="GI:5869938"
/translation="IGAGLGHIGIQVLGALTAETIVVDNRNPDVLDLAVSGADHGV
LADGHVDRVRELTGCGHGAEDVDFVGGEGTGDCIMLRAGDHYHVGNGENIDVPT
IDVISAFINELGNLGSYNDLCELMVLAARGVRLHTAKYSLDRFDQDLDLGGRRIR
GRALVP"
1. .505
/gene="SCF1.01"
558. .887
/gene="SCF1.02"
558. .887
/gene="SCF1.02"
/note="SCF1.02, unknown, len: 109aa; similar to TR:006569
(EMBL:295385) hypothetical protein from Mycobacterium
tuberculosis (107 aa) fasta scores; opt: 433, z-score:
595.2, E(): 8.7e-26, (56.6% identity in 106 aa overlap)."
/codon_start=1
/transl_table=11
/label=SCF1.02
/product="hypothetical protein"
/protein_id="CAB55522.1"
/db_xref="GI:5869939"
/translation="MIFITAFRVLPPEHADQWPOVEDFTRATRAEPGCLWFQWSRLS
DDPTEYVLVEAFRDEAGAAHVGSAAHKAQAQTLPPHLLATTPRIVNANVPQDDWSLLG
EMAVDAG"
complement(967. .1494)
/gene="SCF1.03c"
complement(967. .1494)
/gene="SCF1.03c"
/note="SCF1.03c, probable acetyltransferase, len: 175aa;
similar to many eg. SW:MAA_ECOLI maltose
0-acetyltransferase from Escherichia coli (182 aa) fasta
scores; opt: 365, z-score: 432.1, E(): 1.1e-16, (43.6%
identity in 140 aa overlap). Contains Pfam match to entry
PF00132 hexapep, Bacterial transferase hexapeptide (four
repeats) and Prosite match to PS00101 Hexapeptide-repeat
containing-transferases signature."
/codon_start=1
/transl_table=11
/label=SCF1.03c
/product="putative acetyltransferase"
/protein_id="CAB55523.1"
/db_xref="GI:5869940"
/translation="WVHTPEFARHAERIVEVTDATSLNVLDPFSDSAGRSELLSVVF
GGPLPESVTIYPPFTTXGLNTFCGVFNVOGCTFMDKGGIRGNRYMIAPKASLVT
GGHPLPLARRAHLSEFPIVEDDVGWIGTAAVITQGVIGAGVAVAGAVVTRDVPAG
TVVAGVPARVHKQIG"
complement(1024. .1128)
/gene="SCF1.03c"
/note="Pfam match to entry PF00132 hexapep, Bacterial
transferase hexapeptide (four repeats), score 49.50,
E-value 3.2e-12"
complement(1024. .1110)
/gene="SCF1.03c"
/note="PS00101 Hexapeptide-repeat containing-transferases
signature."
1608. .2543
/gene="SCF1.04"
1608. .2543
/gene="SCF1.04"
/note="SCF1.04, probable LysR-family transcriptional
regulator, len: 311aa; similar to many transcriptional
regulators especially in the N-terminal region eg.
SW:OXYR_MYCAV regulator of hydrogen peroxide responsive
gene from Mycobacterium avium (311 aa) fasta scores; opt:
292, z-score: 329.4, E(): 5.5e-11, (36.6% identity in 191
aa overlap). Contains Pfam match to entry PF00126 HTH_1,

```

```

Bacterial regulatory helix-turn-helix protein, lysR family
and prosite match to PS00044 Bacterial regulatory
proteins, lysR family signature."
/codon_start=1
/transl_table=11
/label=SCF1.04
/product="putative LysR-family transcriptional regulator"
/protein_id="CAB55524.1"
/db_xref="GI:5869941"
/translation="MDVEALRTFVVAGVQFAAADLGLISQOAVSKRIAALRRHVA
VTLVRSRGRSLSDGGVFLPHAKKLVLTVEQAEAVNPGSRPLRVDVNLNRRISPAQ
AVFYRSRPEPTDLDVLDKRENAQAQAVLEGVVDASFALPADRVPTGITAERLL
DEPLELLYPGHPLADAPRISEVDLAGRIWIPGIRCTEWAATFQALSEAFGLSIDA
LCPNFGDEALMDALDSASLATLVGRGDRYLWPRTHDLRRLPLHDPFPFYPHLLLFRT
GDHPVLTAALRDHLRTTAPRTPHDAWTPDWTVTWH"
1614. .2033
/gene="SCF1.04"
/note="Pfam match to entry PF00126 HTH_1, Bacterial
regulatory helix-turn-helix protein, lysR family, score
83.50, E-value 4.3e-21"
1656. .1748
/gene="SCF1.04"
/note="PS00044 Bacterial regulatory proteins, lysR family
signature."
complement(2567. .3682)
/gene="SCF1.05c"
complement(2567. .3682)
/gene="SCF1.05c"
/note="SCF1.05c, possible oxidoreductase, len: 371aa;
weakly similar to many eg. TR:069945 (EMBL:AL023862)
putative oxidoreductase from Streptomyces coelicolor (430
aa) fasta scores; opt: 224, z-score: 247.7, E(): 2e-06, to
(26.0% identity in 392 aa overlap). Contains Pfam match to
entry PF01408 GFO_IDH_MocA, Oxidoreductase family."
/codon_start=1
/transl_table=11
/label=SCF1.05c
/product="putative oxidoreductase"
/protein_id="CAB55525.1"
/db_xref="GI:5869942"
/translation="MNTPGSALADLARSGRPIGVIGLGAHGSWAERSHLPALRAVG
GFELRALSTSSKSGKHGVRRAFGTAELACDEVDLVVAVKVPHRELVRT
VLNACKSVLCENPLGNLAFAEDMARLARAAVPTVVGLOARSHFAPAVRHLVADGY
VGEVLSITVVGSGGANGAVPADHVLDAANGATLLTVFGHLDGASVLGPEDEL
RIOAASRTSADVDTETGRVPVMTAQQVVASGELPGAVATPHYGSGMGTNFRWEI
NGTGDVLTPAPVGHFQPLSLTLEGGRTSLALAPITVEAYARVARLEPRADAPAYA
VAHAYQRFLLDLREGTAHVDFAHGVVHRHSIEQYLP"

```

```

Query Match      81.7%  Score 18.8;  DB 3;  Length 14750;
Best Local Similarity  90.98;  Pred. No. 5.2e+02;
Matches 20;  Conservative  0;  Mismatches 2;  Indels  0;  Gaps  0;

Oy  2  tgcacagccgcaaggcctggga 23
      ||| ||||| ||||| ||||| |||
Db  2253  TGCTCAGGCCGAGGCTCGGA 2232

RESULT 13
LOCUS  AP001078
DEFINITION  Homo sapiens chromosome 18 clone RP11-861B19 map 18p11.3, WORKING
DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION  AP001078
VERSION  AP001078.2  GI:8117771
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE  Homo sapiens
ORGANISM  Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 176253)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,N., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE  Homo sapiens 176,253 genomic DNA of 18p11.3

```



JOURNAL  
REFERENCE  
AUTHORS

Published Only in Database (2000) In press  
2 (bases 1 to 176253)

Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Totsuki, Y., Watanabe, H., and Sakaki, Y.

TITLE  
JOURNAL

Submitted (17-JAN-2000) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
Japan (E-mail: hattori@gsc.riken.go.jp)  
URL: (http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,  
Fax: 81-42-778-9924)

## COMMENT

On May 31, 2000 this sequence version replaced gi:6997425.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-861B19

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 155758 bases at least Q40

Consensus quality: 165815 bases at least Q30

Consensus quality: 170532 bases at least Q20

Insert size: 173753; sum-of-contigs

Quality coverage: 4.13x in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently consists of  
26 contigs. The true order of the pieces is not known and their  
order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved

```

1      25481 contig of 25481 bp in length
25582 44141 contig of 18560 bp in length
44242 59207 contig of 14966 bp in length
59308 73163 contig of 13856 bp in length
73264 83431 contig of 10168 bp in length
83532 90995 contig of 7464 bp in length
91096 99746 contig of 8651 bp in length
99847 105997 contig of 6151 bp in length
106098 112950 contig of 5024 bp in length
113051 118074 contig of 5838 bp in length
118175 124012 contig of 5257 bp in length
124113 129369 contig of 5641 bp in length
129470 135110 contig of 5936 bp in length
135211 141146 contig of 5364 bp in length
141247 146610 contig of 4112 bp in length
146711 150822 contig of 4308 bp in length
150923 155230 contig of 3430 bp in length
155331 158760 contig of 2446 bp in length
158861 163143 contig of 4283 bp in length
163244 165114 contig of 1871 bp in length
165215 167768 contig of 2554 bp in length
167869 170314 contig of 100 bp in length
170415 170414 contig of 100 bp in length
170415 171946 contig of 1532 bp in length
171947 172046 contig of 100 bp in length
172047 173708 contig of 1662 bp in length
173709 173808 contig of 100 bp in length
173809 174921 contig of 1113 bp in length
174922 175021 contig of 100 bp in length
175022 176253 contig of 1232 bp in length
Sequence updated (26-May-2000).

```

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 25481: contig of 25481 bp in length

```

*      25482 25581: gap of 100 bp
25582 44141: contig of 18560 bp in length
44142 44241: gap of 100 bp
44242 59207: contig of 14966 bp in length
59208 59307: gap of 100 bp
59308 73163: contig of 13856 bp in length
73164 73263: gap of 100 bp
73264 83431: contig of 10168 bp in length
83432 83531: gap of 100 bp
83532 90995: contig of 7464 bp in length
90996 91095: gap of 100 bp
91096 99746: contig of 8651 bp in length
99747 99846: gap of 100 bp
99847 105997: contig of 6151 bp in length
105998 106097: gap of 100 bp
106098 112950: contig of 6853 bp in length
112951 113050: gap of 100 bp
113051 118074: contig of 5024 bp in length
118075 118174: gap of 100 bp
118175 124012: contig of 5838 bp in length
124013 124112: gap of 100 bp
124113 129369: contig of 5257 bp in length
129370 129469: gap of 100 bp
129470 135110: contig of 5641 bp in length
135111 135210: gap of 100 bp
135211 141146: contig of 5936 bp in length
141147 141246: gap of 100 bp
141247 146610: contig of 5364 bp in length
146611 146710: gap of 100 bp
146711 150822: contig of 4112 bp in length
150823 150922: gap of 100 bp
150923 155230: contig of 4308 bp in length
155231 155330: gap of 100 bp
155331 158760: contig of 3430 bp in length
158761 158860: gap of 100 bp
158861 163143: contig of 4283 bp in length
163144 163243: gap of 100 bp
163244 165114: contig of 1871 bp in length
165115 165214: gap of 100 bp
165215 167768: contig of 2554 bp in length
167769 167868: gap of 100 bp
167869 170314: contig of 2446 bp in length
170315 170414: gap of 100 bp
170415 171946: contig of 1532 bp in length
171947 172046: gap of 100 bp
172047 173708: contig of 1662 bp in length
173709 173808: gap of 100 bp
173809 174921: contig of 1113 bp in length
174922 175021: gap of 100 bp
175022 176253: contig of 1232 bp in length.

```

FEATURES  
Source

```

1..176253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-861B19"

```

```

misc_feature
1..25481
/note="assembly_fragment"
misc_feature
25582..44141
/note="assembly_fragment"
misc_feature
44242..59207
/note="assembly_fragment"
misc_feature
59308..73163
/note="assembly_fragment"
misc_feature
73264..83431
/note="assembly_fragment"
misc_feature
83532..90995
/note="assembly_fragment"
misc_feature
91096..99746
/note="assembly_fragment"
misc_feature
99847..105997
/note="assembly_fragment"

```

COMMENT

|           |        |        |        |    |       |    |    |        |
|-----------|--------|--------|--------|----|-------|----|----|--------|
| preserved | 1      | 15033  | contig | of | 15033 | bp | in | length |
|           | 15134  | 28850  | contig | of | 13717 | bp | in | length |
|           | 28951  | 43381  | contig | of | 14431 | bp | in | length |
|           | 43482  | 54311  | contig | of | 10830 | bp | in | length |
|           | 54412  | 65508  | contig | of | 11097 | bp | in | length |
|           | 65609  | 74474  | contig | of | 8866  | bp | in | length |
|           | 74575  | 86282  | contig | of | 11708 | bp | in | length |
|           | 86383  | 94094  | contig | of | 8712  | bp | in | length |
|           | 94195  | 102440 | contig | of | 7246  | bp | in | length |
|           | 102541 | 109661 | contig | of | 7121  | bp | in | length |
|           | 109762 | 118186 | contig | of | 8425  | bp | in | length |
|           | 118287 | 123309 | contig | of | 5023  | bp | in | length |
|           | 123410 | 138976 | contig | of | 5567  | bp | in | length |
|           | 123417 | 133984 | contig | of | 4908  | bp | in | length |
|           | 129077 | 138474 | contig | of | 4390  | bp | in | length |
|           | 134085 | 143774 | contig | of | 5200  | bp | in | length |
|           | 138575 | 143774 | contig | of | 5795  | bp | in | length |
|           | 143875 | 149669 | contig | of | 3767  | bp | in | length |
|           | 149770 | 153536 | contig | of | 4387  | bp | in | length |
|           | 153637 | 158023 | contig | of | 4322  | bp | in | length |
|           | 158124 | 162445 | contig | of | 2698  | bp | in | length |
|           | 162546 | 165243 | contig | of | 3532  | bp | in | length |
|           | 165344 | 168875 | contig | of | 2998  | bp | in | length |
|           | 168976 | 171973 | contig | of | 3916  | bp | in | length |
|           | 172074 | 175989 | contig | of | 3692  | bp | in | length |
|           | 176090 | 179781 | contig | of | 2362  | bp | in | length |
|           | 179882 | 182243 | contig | of | 2222  | bp | in | length |
|           | 182344 | 184565 | contig | of | 2482  | bp | in | length |
|           | 184666 | 187147 | contig | of | 2301  | bp | in | length |
|           | 187248 | 189548 | contig | of | 1013  | bp | in | length |
|           | 189649 | 190661 | contig | of | 1206  | bp | in | length |
|           | 190762 | 191967 | contig | of | 1420  | bp | in | length |
|           | 192068 | 193487 | contig | of |       |    |    |        |

Sequence updated (26-May-2000).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 32 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* be preserved.  
 \* 1 15033: contig of 15033 bp in length  
 \* 15034 15133: gap of 100 bp  
 \* 15134 28850: contig of 13717 bp in length



Thu Oct 25 13:09:01 2001

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 8, 2000 this sequence version replaced gi:10047674.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L1313  
Center clone name: 168\_M\_18  
----- Summary Statistics  
Sequencing vector: M13; M77815; 19% of reads  
Sequencing vector: Plasmid; n/a; 81% of reads  
Chemistry: Dye-terminator; Big Dye; 97% of reads  
Chemistry: Dye-terminator; Big Dye; 97% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 195941 bases at least Q40  
Consensus quality: 202775 bases at least Q30  
Consensus quality: 206217 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 208534; sum-of-contigs  
Quality coverage: 24.4 in Q20 base.  
NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 26180: contig of 26180 bp in length  
26181 26280: gap of 100 bp  
26281 27483: contig of 1203 bp in length  
27484 27583: gap of 100 bp  
27584 28811: contig of 1228 bp in length  
28812 28911: gap of 100 bp  
28912 30054: contig of 1143 bp in length  
30055 30154: gap of 100 bp  
30155 31369: contig of 1215 bp in length  
31370 31469: gap of 100 bp  
31470 32669: contig of 1200 bp in length  
32670 32769: gap of 100 bp  
32770 33804: contig of 1035 bp in length  
33805 33904: gap of 100 bp  
33905 35717: contig of 1813 bp in length  
35718 35817: gap of 100 bp  
35818 37338: contig of 1521 bp in length  
37339 37438: gap of 100 bp  
37439 39081: contig of 1643 bp in length  
39082 39181: gap of 100 bp  
39182 40202: contig of 1021 bp in length  
40203 40302: gap of 100 bp  
40303 41938: contig of 1636 bp in length  
41939 42038: gap of 100 bp  
42039 43493: contig of 1455 bp in length  
43494 43593: gap of 100 bp  
43594 44979: contig of 1386 bp in length  
44980 45079: gap of 100 bp  
45080 46806: contig of 1727 bp in length  
46807 46906: gap of 100 bp  
46907 48423: contig of 1517 bp in length  
48424 48523: gap of 100 bp  
48524 49581: contig of 1058 bp in length  
49582 49681: gap of 100 bp  
49682 51224: contig of 1543 bp in length  
51225 51324: gap of 100 bp  
51325 52909: contig of 1595 bp in length  
52910 53009: gap of 100 bp

\* 53010 54846: contig of 1837 bp in length  
\* 54847 54946: gap of 100 bp  
\* 54947 56941: contig of 1995 bp in length  
\* 56942 57041: gap of 100 bp  
\* 57042 60058: contig of 3017 bp in length  
\* 60059 60158: gap of 100 bp  
\* 60159 210734: contig of 150576 bp in length.  
FEATURES  
Location/Qualifiers  
1. .210734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18"  
/clone="RP11-168M18"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .26180  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
26281..27483  
/note="assembly\_fragment"  
27584..28811  
/note="assembly\_fragment"  
28912..30054  
/note="assembly\_fragment"  
30155..31369  
/note="assembly\_fragment"  
31470..32669  
/note="assembly\_fragment"  
32770..33804  
/note="assembly\_fragment"  
33905..35717  
/note="assembly\_fragment"  
35818..37338  
/note="assembly\_fragment"  
37439..39081  
/note="assembly\_fragment"  
39182..40202  
/note="assembly\_fragment"  
40303..41938  
/note="assembly\_fragment"  
42039..43493  
/note="assembly\_fragment"  
43594..44979  
/note="assembly\_fragment"  
45080..46806  
/note="assembly\_fragment"  
46907..48423  
/note="assembly\_fragment"  
48524..49581  
/note="assembly\_fragment"  
49682..51224  
/note="assembly\_fragment"  
51325..52909  
/note="assembly\_fragment"  
53010..54846  
/note="assembly\_fragment"  
54947..56941  
/note="assembly\_fragment"  
57042..60058  
/note="assembly\_fragment"  
60159..210734  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
BASE COUNT 57661 a 45068 c 43811 g 61994 t 2200 others  
ORIGIN

Query Match 81.7%; Score 18.8; DB 63; Length 210734;  
Best Local Similarity 90.9%; Pred No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

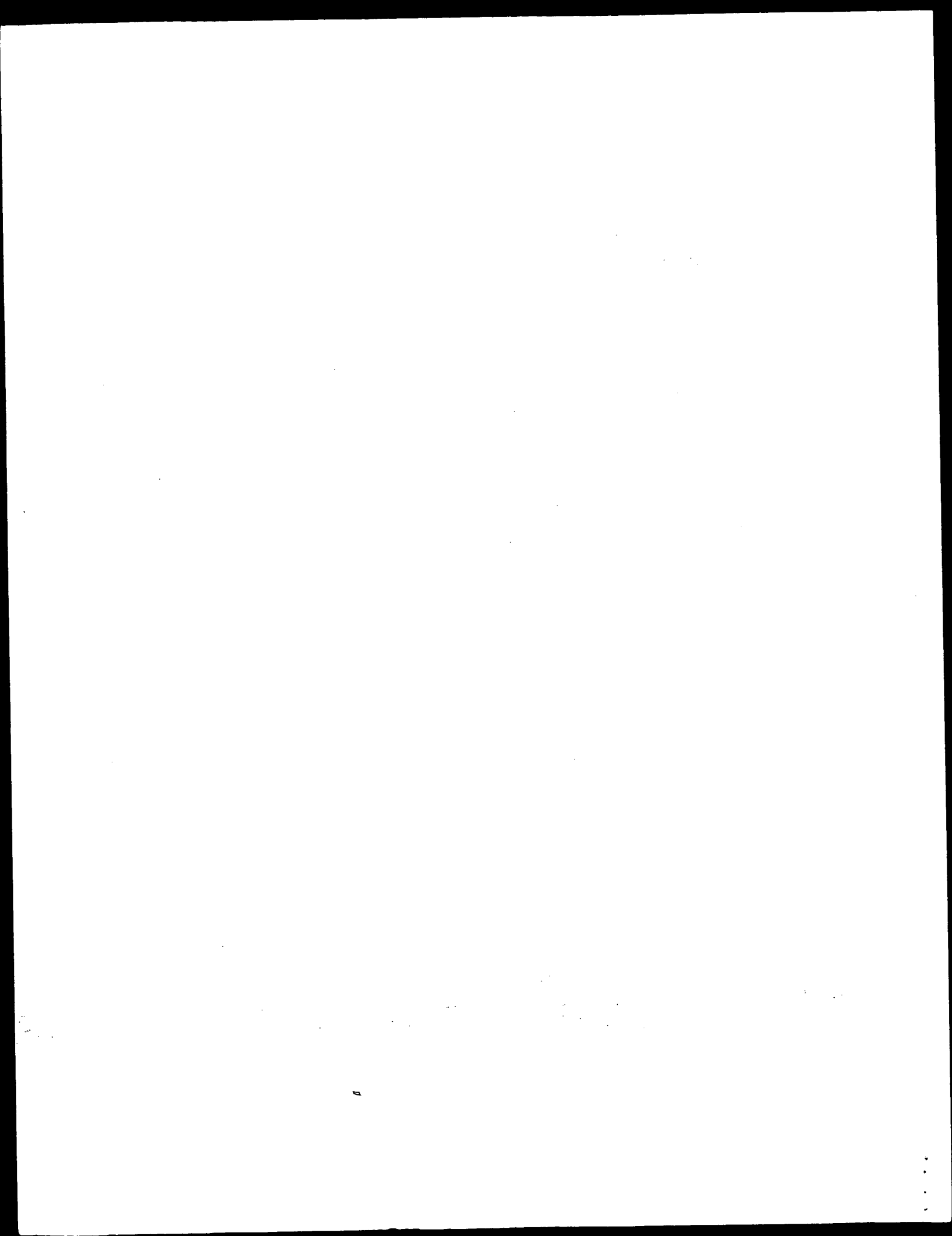
Thu Oct 25 13:09:01 2001

us-09-462-955-6.rge

Page 17

Qy 1 ctgcccaggccgaaggcctggg 22  
||||| ||||| ||||| |||||  
Db 203306 CTGCCAAGGCCCAAGGCCTGGG 203285

Search completed: October 24, 2001, 11:43:17  
Job time: 6382 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:31 : Search time 180.6 Seconds  
(without alignments)  
24.109 Million cell updates/sec

Title: US-09-462-955-6

Perfect score: 23  
Sequence: 1 ctgcccagccggaagcctggga 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description      |
|------------|-------|-------------|--------|-------|------------------|
| 1          | 18.2  | 79.1        | 3131   | 3     | US-09-035-648-23 |
| 2          | 18.2  | 79.1        | 3131   | 4     | US-09-001-951-23 |
| 3          | 17.8  | 77.4        | 2816   | 1     | US-08-785-241-1  |
| 4          | 16.6  | 72.2        | 1659   | 3     | US-09-083-351-3  |
| 5          | 16.6  | 72.2        | 1659   | 4     | US-09-083-352-3  |
| 6          | 16.6  | 72.2        | 2828   | 3     | US-09-357-071-1  |
| 7          | 16.6  | 72.2        | 3946   | 3     | US-09-083-351-1  |
| 8          | 16.6  | 72.2        | 3946   | 4     | US-09-083-352-1  |
| 9          | 16.4  | 71.3        | 4200   | 1     | US-07-841-654B-1 |
| 10         | 16.4  | 71.3        | 4200   | 1     | US-07-946-234A-1 |
| 11         | 16.4  | 71.3        | 4200   | 1     | US-08-123-161A-1 |
| 12         | 16.4  | 71.3        | 4200   | 1     | US-08-483-278-1  |
| 13         | 16.4  | 71.3        | 4200   | 5     | PCT-US93-01560-1 |
| 14         | 16.2  | 70.4        | 472    | 2     | US-08-811-949-40 |
| 15         | 16.2  | 70.4        | 747    | 1     | US-08-257-341-6  |
| 16         | 16.2  | 70.4        | 779    | 1     | US-08-133-804-3  |
| 17         | 16.2  | 70.4        | 779    | 1     | US-08-461-838-3  |
| 18         | 16.2  | 70.4        | 779    | 2     | US-08-461-838-3  |
| 19         | 16.2  | 70.4        | 942    | 3     | US-08-732-412-1  |
| 20         | 16.2  | 70.4        | 1065   | 1     | US-08-427-640-1  |
| 21         | 16.2  | 70.4        | 1065   | 1     | US-08-427-640-5  |
| 22         | 16.2  | 70.4        | 1065   | 2     | US-08-811-949-60 |
| 23         | 16.2  | 70.4        | 1068   | 1     | US-08-137-116-2  |
| 24         | 16.2  | 70.4        | 1068   | 1     | US-08-427-640-3  |
| 25         | 16.2  | 70.4        | 1068   | 2     | US-08-811-949-44 |
| 26         | 16.2  | 70.4        | 1068   | 2     | US-08-811-949-46 |
| 27         | 16.2  | 70.4        | 1068   | 2     | US-08-811-949-52 |

|      |      |      |      |   |                   |
|------|------|------|------|---|-------------------|
| C 28 | 16.2 | 70.4 | 1068 | 2 | US-08-811-949-58  |
| C 29 | 16.2 | 70.4 | 1068 | 6 | 5223256-3         |
| C 30 | 16.2 | 70.4 | 1110 | 1 | US-08-257-341-4   |
| C 31 | 16.2 | 70.4 | 1170 | 2 | US-08-811-949-64  |
| C 32 | 16.2 | 70.4 | 1170 | 2 | US-08-811-949-66  |
| C 33 | 16.2 | 70.4 | 1260 | 1 | US-08-029-404-1   |
| C 34 | 16.2 | 70.4 | 1260 | 3 | US-08-459-953A-1  |
| C 35 | 16.2 | 70.4 | 1314 | 2 | US-08-811-949-50  |
| C 36 | 16.2 | 70.4 | 1314 | 2 | US-08-811-949-50  |
| C 37 | 16.2 | 70.4 | 1314 | 2 | US-08-811-949-54  |
| C 38 | 16.2 | 70.4 | 1314 | 2 | US-08-811-949-56  |
| C 39 | 16.2 | 70.4 | 1419 | 2 | US-08-811-949-62  |
| C 40 | 16.2 | 70.4 | 1461 | 1 | US-08-403-634-1   |
| C 41 | 16.2 | 70.4 | 1461 | 4 | US-08-913-441B-1  |
| C 42 | 16.2 | 70.4 | 1724 | 6 | 5200340-5         |
| C 43 | 16.2 | 70.4 | 1738 | 6 | 5200340-1         |
| C 44 | 16.2 | 70.4 | 1848 | 3 | US-08-814-412-10  |
| C 45 | 16.2 | 70.4 | 1955 | 2 | US-08-883-795A-39 |

## ALIGNMENTS

## RESULT 1

US-09-035-648-23  
; Sequence 23, Application US/09035648  
; Patent No. 6100031  
; GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL

TITLE OF INVENTION: GROWTH AND PROLIFERATION

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/035,648

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/818,829

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Meiklejohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283

REFERENCE/DOCKET NUMBER: 07334/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 3131 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 449...2665

US-09-035-648-23

RESULT 3  
S-08-785-241-1/c  
Sequence 1. Application US/08785241

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1659 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-083-351-3

```

```

Query Match          72.2%; Score 16.6; DB 3; Length 1659;
Best Local Similarity 82.6%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 ctgcccaggccgaagcctggga 23
||||| 11 |||||
Db 743 CTGCCAGGGCGGGCGGGGA 721

```

```

RESULT 5
US-09-083-352-3/c
; Sequence 3, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1659 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-083-352-3

Query Match          72.2%; Score 16.6; DB 4; Length 1659;
Best Local Similarity 82.6%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctgcccaggccgaagcctggga 23
||||| 11 |||||
Db 743 CTGCCAGGGCGGGCGGGGA 721

RESULT 6
US-09-357-071-1
; Sequence 1, Application US/09357071
; Patent No. 6043091
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF LIVER GLYCOGEN PHOSPHORYLASE EXPRESSION
; FILE REFERENCE: RTS-0074
; CURRENT APPLICATION NUMBER: US/09/357,071
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2828
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (114)..(2657)
US-09-357-071-1

```

```

Query Match          72.2%; Score 16.6; DB 3; Length 2828;
Best Local Similarity 82.6%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctgcccaggccgaagcctggga 23
||||| 11 |||||
Db 1191 ctgcccaggccgaagcctggga 1213

```

```

RESULT 7
US-09-083-351-1/c
; Sequence 1, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351

```

```
;
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 475..2133
; US-09-083-351-1

Query Match 72.2%; Score 16.6; DB 3; Length 3946;
Best Local Similarity 82.6%; Pred. No. 87;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctgcccagggcgaagcctggga 23
||||| || ||| |||
Db 1217 CTGCCAGGGCGCGCGGGGA 1195

RESULT 9
US-07-841-654B-1/c
; Sequence 1, Application US/07841654B
; Patent No. 5260209
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Ibraghimov-Beskrovnya, Oxana
; APPLICANT: Ervasti, James M.
; APPLICANT: Leveille, Cynthia J.
; APPLICANT: Matsumura, Kiichiro
; TITLE OF INVENTION: DNA ENCODING DYSTROPHIN-ASSOCIATED
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,654B
; FILING DATE: 19920220
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: UIR89-11AA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 861-6240
; TELEFAX: 617 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
; US-07-841-654B-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0;

QY 6 caggccgaagcctggga 23
||||| ||| ||| |||
Db 4051 CAGGCCCAAGCCTGGGA 4034
```

RESULT 10  
US-07-946-234A-1/c  
; Sequence 1, Application US/07946234A  
; Patent No. 5308752  
; GENERAL INFORMATION:  
; APPLICANT: Campbell, Kevin P.  
; TITLE OF INVENTION: DIAGNOSIS OF AUTOSOMAL MUSCULAR DYSTROPHY  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS  
; STREET: 2 Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/946,234A  
; FILING DATE: 19920914  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brook, David E.  
; REGISTRATION NUMBER: 22,592  
; REFERENCE/DOCKET NUMBER: UIRF89-11AAA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)861-6240  
; TELEFAX: (617)861-9540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4200 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 170..2855  
US-07-946-234A-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;  
Best Local Similarity 94.4%; Pred. No. 1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 caggccgaagcctggga 23  
||||| |||||||  
Db 4051 CAGGCCCAAGCCTGGGA 4034

RESULT 11  
US-08-123-161A-1/c  
; Sequence 1, Application US/08123161A  
; Patent No. 5449616  
; GENERAL INFORMATION:  
; APPLICANT: Campbell, Kevin P.  
; APPLICANT: Roberts, Steven L.  
; APPLICANT: Anderson, Richard D.  
; APPLICANT: Ibraghimov, Oxana B.  
; APPLICANT: Yang, Bin  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P.C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME

; COUNTRY: USA  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/123,161A  
; FILING DATE: 16-SEP-93  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/946,234  
; FILING DATE: 14-SEP-92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: UIRF89-11A4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (207) 363-0558  
; TELEFAX: (207) 363-0528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 170..2855  
US-08-123-161A-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;  
Best Local Similarity 94.4%; Pred. No. 1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 caggccgaagcctggga 23  
||||| |||||||  
Db 4051 CAGGCCCAAGCCTGGGA 4034

RESULT 12  
US-08-483-278-1/c  
; Sequence 1, Application US/08483278  
; Patent No. 5686073  
; GENERAL INFORMATION:  
; APPLICANT: Campbell, Kevin P.  
; APPLICANT: Ibraghimov, Oxana B.  
; APPLICANT: Ervasti, James M.  
; APPLICANT: Leveille, Cynthia J.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell, P.C.  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: ME  
; COUNTRY: USA  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,278  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/123,161

```
;
; FILING DATE: 16-SEP-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIR89-11A5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
;
US-08-483-278-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 caggccgaaggcctggga 23
||||| |||||||
Db 4051 CAGGCCCAAGGCTGGGA 4034

RESULT 13
PCT-US93-01560-1/c
; Sequence 1, Application PC/TUS9301560
; GENERAL INFORMATION:
; APPLICANT: University of Iowa Research Foundation
; TITLE OF INVENTION: DNA ENCODING DYSTROPHIN-ASSOCIATED
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/01560
FILING DATE: 19930219
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/841,654
FILING DATE: 20-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: UIR89-11AA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 861-6240
TELEFAX: 617 861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4200 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS

;
; FILING DATE: 16-SEP-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIR89-11A5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
;
US-08-483-278-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 caggccgaaggcctggga 23
||||| |||||||
Db 4051 CAGGCCCAAGGCTGGGA 4034

RESULT 14
US-08-811-949-40/c
; Sequence 40, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,949
FILING DATE: 05-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-966-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..471
;
US-08-811-949-40

Query Match 70.4%; Score 16.2; DB 2; Length 472;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgcccagccgaaggcctggg 22
||||| |||||||
Db 70 TGCCCAAGGCCAGTGGCTGGG 50

RESULT 15
```

US-08-257-341-6  
 ; Sequence 6, Application US/08257341  
 ; Patent No. 5525491  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HUSTON, JAMES S  
 ; APPLICANT: OPPERMAN, HERMANN  
 ; APPLICANT: TIMASHEFF, SERGE N  
 ; TITLE OF INVENTION: SERINE RICH PEPTIDE LINKER  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CREATIVE BIOMOLECULES, INC./PATENT DEPT.  
 ; STREET: 35 SOUTH STREET  
 ; CITY: HOPKINTON  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 01748  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/257,341  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/842,149  
 ; FILING DATE:  
 ; APPLICATION NUMBER: US 07/662,226  
 ; FILING DATE: 27-FEB-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CAMPBELL ESQ, PAULA A  
 ; REGISTRATION NUMBER: 32,503  
 ; REFERENCE/DOCKET NUMBER: CRP-064CP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/248-7000 (ATTY)  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 747 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: syn DNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..747  
 ; US-08-257-341-6

Query Match 70.4%; Score 16.2; DB 1; Length 747;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 tgcccagccgaagcctggg 22  
 ||| |  
 Db 645 TGTCGAGCCGAGAGACTGGG 665

Search completed: October 24, 2001, 10:00:32  
 Job time: 217 sec

